

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**SOL ATRİYUM ÇAPININ TEMEL BİLEŞENLER REGRESYONU,
KİSMİ EN KÜÇÜK KARELER REGRESYONU ve YAPAY SİNİR
AĞLARI İLE TAHMİNLERİNİN KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

Fatma AŞKIN

Anabilim Dalı: İstatistik
Programı: Uygulamalı İstatistik

Tez Danışmanı: Yrd. Doç. Dr. Cemil ÇOLAK

ELAZIĞ-2011

**T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**SOL ATRİYUM ÇAPININ TEMEL BİLEŞENLER REGRESYONU, KISMİ EN
KÜÇÜK KARELER REGRESYONU ve YAPAY SİNİR AĞLARI İLE
TAHMİNLERİNİN KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

**Fatma AŞKIN
(091133107)**

Tezin Enstitüye Verildiği Tarih: 13.07.2011

Tezin Savunulduğu Tarih: 28.07.2011

Tez Danışmanı: Yrd. Doç. Dr. Cemil ÇOLAK (İ.Ü)

**Diğer Jüri Üyeleri: Yrd. Doç. Dr. Sinan ÇALIK (F.Ü)
Yrd. Doç. Dr. Reşat YILMAZER (F.Ü)**

TEMMUZ-2011

ÖNSÖZ

Tez konusunun belirlenmesi ve yürütülmesi aşamasında, her türlü yardımı ve desteği esirgemeyen başta danışman hocam Yrd. Doç. Dr. Cemil ÇOLAK' a, ve Arş. Gör. Esra Pamukçu'ya diğer bilimsel aşamalarda bilgilerini benimle paylaşan değerli hocalarım, Yrd. Doç. Dr.Mahmut IŞIK, Yrd. Doç. Dr. Mehmet GÜRCAN'a, Yrd. Doç. Dr. Sinan ÇALIK'a, Yrd. Doç. Dr. Nurhan HALİSDEMİR'e ve uygulama aşamasında verileri temin ettiğimiz Fırat Üniversitesi Tıp Fakültesi, Kardiyoloji Bölümü öğretim üyesi Doç. Dr. Necati Dağlı ve asistanı Dr. Oğuzkaan Kaya ve tezi hazırlama aşamasında bana destek olan değerli arkadaşım Mehmet Onur Kaya'ya ve abim Fatih Aşkın'a teşekkür ederim.

Fatma AŞKIN
ELAZIĞ-2011

İÇİNDEKİLER

	Sayfa No
ÖNSÖZ.....	II
İÇİNDEKİLER.....	III
ÖZET.....	V
SUMMARY.....	VI
ŞEKİL LİSTESİ.....	VII
TABLolar LİSTESİ.....	VIII
KISALTMALAR LİSTESİ.....	IX
1. GİRİŞ.....	1
2. MATERYAL VE METOT.....	2
2.1. Çoklu Doğrusal Regresyon Modeli.....	2
2.1.1. Çoklu Doğrusal Regresyon Modelinde En Küçük Kareler Metodu İle Parametre Tahmini.....	3
2.1.2. En Küçük Kareler Metodunun Varsayımları.....	4
2.2. Çoklu Doğrusal Bağlantı.....	5
2.2.1. Çoklu Doğrusal Bağlantının Kaynakları.....	6
2.2.2. Çoklu Doğrusal Bağlantının Saptanması.....	7
2.2.3. Çoklu Bağlantı Probleminin Sonuçları.....	7
2.2.4. Çoklu Bağlantı Probleminin Çözümü.....	8
2.3. Temel Bileşenler Analizi.....	8
2.3.1. Temel Bileşenlerin Elde Edilmesi.....	8
2.3.2. Temel Bileşenlerin Özellikleri.....	14
2.3.3. Temel Bileşen Sayısının Belirlenmesi.....	14
2.4. Temel Bileşenler Regresyonu.....	15
2.4.1. Teorik Model.....	16
2.5. Kısmi En Küçük Kareler Regresyonu.....	17
2.5.1. NIPALS Algoritması.....	17
2.6. Yapay Sinir Ağları.....	19
2.6.1. Yapay Sinir Ağlarının Yapısı.....	19

3.	UYGULAMA.....	23
3.1.	Gereç ve Yöntem.....	23
3.1.1.	Araştırmanın Tasarımı ve Modeli	23
3.2.	Bulgular	24
4.	SONUÇ	43
5.	EKLER	44
	EK Tablo 1.....	45
6.	KAYNAKLAR	52
7.	ÖZGEÇMİŞ.....	54

ÖZET

Bu çalışmada sol atriyum çapının, Temel Bileşenler Regresyonu, Kısmi En Küçük Kareler regresyonu ve Yapay Sinir Ağları ile tahmininin yapılması amaçlanmıştır. Bu amaçla ilk önce çoklu doğrusal regresyon analizinde parametre tahmini için yaygın olarak kullanılan En Küçük kareler metodu ve varsayımları kısaca açıklanmıştır. Ardından açıklayıcı değişkenler arasındaki bağımsızlık varsayımının bozulması durumunda oluşan çoklu bağlantı problemi incelenmiştir. Bu problemin giderilmesi için kullanılan Temel Bileşenler Regresyonu ve Kısmi En küçük Kareler Regresyonu açıklanmıştır. Son olarak Yapay Sinir Ağları açıklanmıştır. Uygulama aşamasında, Fırat Üniversitesi Tıp Fakültesi Kardiyoloji Polikliniğine gelen 127 hipertansif hastanın ekokardiyografi raporları prospektif olarak toplandı. Elde edilen verilerle yukarıda açıklanan yöntemlerin tümünü analiz edildi ve sonuçlar karşılaştırıldı.

Anahtar Kelimeler: En Küçük Kareler Metodu, Çoklu Bağlantı Problemi, Temel Bileşenler Regresyonu, Kısmi En Küçük Kareler Regresyonu, Yapay Sinir Ağları.

SUMMARY

COMPARISON OF THE LEFT ATRIUM DIAMETER ESTIMATIONS WITH PRINCIPAL COMPONENTS REGRESSION, PARTIAL LEAST SQUARES REGRESSION AND ARTIFICIAL NEURAL NETWORKS METODS

In this study, it was aimed to estimate of diameter of the left atrium with Principal Component Regression, Partial Least Square Regression and Artifical Neural Networks. For this purpose, firstly, multiple linear regression analysis, Least square method that is commonly used for parameter estimates and its assumptions were briefly described. Secondly, multicollinearity problem that occur in case of failure of assumption of independence between the explanatory variables was examined. Principal components regression and partial least square regression that used to overcome this problem were described. Finally, artificial neural network was examined. In the part of application, echocardiography reports of 127 hypertensive patients who came to Cardiology Polyclinic of Medicine Faculty of Firat University were collected prospectively. The obtained data were analyzed by all of methods described above and the results were compared.

Key Words: Least-Squares Method, The Multicollinearity Problem, Principal Components Regression, Partial Least Squares Regression, Artificial Neural Networks.

ŞEKİL LİSTESİ

	Sayfa No
Şekil 1. Özdeğerlerin Varyans Açıklama Oranları.....	15
Şekil 2. Yapay Sinir Ağı Yapısı	20
Şekil 3. Sigmoid Fonksiyon Grafiği	20
Şekil 4. VIF Grafiği	29
Şekil 5. PLS Model Seçim Grafiği	39

TABLULAR LİSTESİ

	Sayfa No
Tablo 3.1. Modeldeki Değişkenler ve Ölçü Birimleri.....	23
Tablo 3.2. Tanımlayıcı istatistikler	24
Tablo 3.3. EKK Çoklu Regresyon Sonuçları.....	25
Tablo 3.4. EKK için ANOVA tablosu	25
Tablo 3.5. Korelasyon Matrisi	26
Tablo 3.6. EKK çoklu bağlantının saptanması	27
Tablo 3.7. Korelasyon Özdeğerleri	28
Tablo 3.8. Standartlaştırılmış Temel Bileşenler Regresyon Katsayıları	30
Tablo 3.9. VIF değerleri	31
Tablo 3.10. PC Analiz Tablosu.....	32
Tablo 3.11. PC=5 için Temel Bileşenler ve EKK'nın Karşılaştırılması	33
Tablo 3.12. Temel Bileşenler Regresyon Katsayıları	34
Tablo 3.13. PC=5 için ANOVA Tablosu	35
Tablo 4.2. Bileşenlere Ait Ağırlık Değerleri	36
Tablo 4.3. Bileşenlere Ait Yük Değerleri.....	37
Tablo 4.4. PLSR Model Seçimi ve Geçerlilik Tablosu.....	40
Tablo 4.5. Regresyon Katsayıları Tablosu	41
Tablo 4.6. PLSR(2) İçin ANOVA Tablosu	41
Tablo 5.1. YSA Modellerine İlişkin Bilgiler ve Sonuçları.....	42
Tablo 6.1. Tahminlerin Karşılaştırılması.....	43

KISALTMALAR LİSTESİ

ANOVA	: Tek Yönlü Varyans Analizi
EKK	: En Küçük Kareler
HKO	: Hata Kareler Ortalaması
HKT	: Hata Kareler Toplamı
Ö.A	: Öğrenme Algoritması
PC	: Principal Component (Temel Bileşen)
PCR	: Principal Component Regression (Temel Bileşenler Regresyonu)
PLSR	: Partial Least Square Regression (Kısmi En Küçük Kareler Regresyonu)
PRESS	: Predictive Residual Sum of Squares (Tahmini Hata Kareler)
RMSE	: Root Mean Square Error (Hata Kareler Ortalamasının Karekökü)
VIF	: Variance Inflation Factors (Varyans Şişirme Faktörü)
VIP	: Variable Importance in the Projection
YSA	: Yapay Sinir Ağları

1. GİRİŞ

Regresyon analizi, bir bağımlı değişken (Y) ile bir veya birden fazla bağımsız değişken (X) arasındaki fonksiyonel ilişkinin incelenmesi, ileri sürülen bir iddianın test edilmesi, izlenecek politikanın tespiti ve geleceğe dönük ön tahmin yapılması amacıyla kullanılır[12]. Bu analiz tekniğinde iki veya daha fazla değişken arasındaki ilişkiyi açıklamak için matematiksel model kullanılır ve bu modele “regresyon modeli” denir. Bağımlı değişkeni açıklamak için bir bağımsız değişken kullanılacaksa basit regresyon, iki veya daha fazla bağımsız değişken kullanılacaksa çoklu regresyon analizi kullanılır. Değişkenler arasında doğrusal ilişki olacağı gibi doğrusal olmayan bir ilişki de olabilir. Basit ve çoklu doğrusal regresyon teknikleri değişkenler arasında doğrusal bir ilişki olduğunda kullanılır[20].

Birçok alanda olduğu gibi sağlık alanında da bir faktörün ortaya çıkışını birden çok faktör etkilemektedir. Örneğin kadınlarda insülin direncini (bağımlı değişken) tahmin etmek için bir regresyon modeli oluşturulmak istenirse, insülin direncine etki eden trigliserit düzeyi dikkate alınarak bu iki değişken arasında basit regresyon modeli oluşturulabilir. Eğer trigliserit düzeyinin insülin direncini açıklamakta yetersiz kaldığı görülürse, insülin direncine etki ettiği düşünülen, HDL kolesterol düzeyi, LDL kolesterol düzeyi, kişinin yaşı gibi birçok faktör (bağımsız değişken) modele dahil edilerek bir çoklu regresyon modeli oluşturulabilir.

Bu çalışmada çoklu doğrusal regresyon kısaca açıklanarak, çoklu doğrusal regresyon analizinde parametre tahmini için en yaygın yöntem olarak kullanılan En Küçük Kareler Metodu ve varsayımları üzerinde durulacaktır. Açıklayıcı değişkenler arasındaki bağımsızlık varsayımının bozulmasıyla ortaya çıkan çoklu bağlantı probleminde değinilerek bu problemin çözüm yöntemlerinden olan Temel Bileşenler Regresyonu ve Kısmi En Küçük Kareler Regresyonu ele alınacaktır. Ayrıca birçok çalışmada çeşitli regresyon yöntemlerinden daha iyi sonuç verdiği görülen [14,23,24] Yapay Sinir Ağları incelenecek, sağlık alanında bir uygulama yapılarak bu yöntemler karşılaştırılacaktır.

2. MATERYAL VE METOT

2.1. Çoklu Doğrusal Regresyon Modeli

Bağımlı değişkeni açıklamak için kullanılan çoklu regresyon modeli,

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i$ şeklinde gösterilir. Burada $i=1,2,\dots,n$ gözlem sayısı, p açıklayıcı (bağımsız) değişken sayısını göstermektedir ($n>p$).

Çoklu doğrusal regresyon modeli matris biçiminde ifade edilirse;

Y ; $n \times 1$ boyutlu bağımlı değişken vektörüdür.

X ; $n \times (p+1)$ boyutlu bağımsız değişkenlerden oluşan matristir.

β ; $(p+1) \times 1$ boyutlu bilinmeyen parametreler vektörünü göstermektedir. β parametreleri Bağımlı değişkendeki bir birimlik değişiminin bağımsız değişkendeki değişimini veren kısmi regresyon katsayılarıdır.

ε ; $n \times 1$ boyutlu normal dağılıma sahip, beklenen değeri sıfır, varyansı (σ^2) sabit, ilişkisiz, bağımsız rassal hata vektörüdür.

$$Y = X\beta + \varepsilon \text{ veya}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1n} \\ 1 & X_{21} & X_{22} & \dots & X_{2n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & X_{p1} & X_{p2} & \dots & X_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix} \text{ şeklinde yazılabilir.}$$

Çoklu doğrusal regresyonda iki temel amaç vardır:

- 1- Bağımlı değişkeni etkilediği düşünülen bağımsız değişkenlerden hangisi ya da hangilerinin bağımlı değişkeni daha çok etkilediğini bulmak.
- 2- Bağımlı değişkeni etkilediği belirlenen bağımsız değişkenler yardımıyla bağımlı değişken değerini kestirebilmek[20].

2.1.1. Çoklu Doğrusal Regresyon Modelinde En Küçük Kareler Metodu İle Parametre Tahmini

En küçük kareler tahmin yöntemi, gözlenmiş Y_j değerleri ile bunların tahmin değerleri arasındaki farkın kareler toplamını, bir başka deyişle hata kareler toplamını minimum yapacak çözümü bulmak için kullanılan bir yöntemdir. $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$; $\beta_0, \beta_1, \dots, \beta_p$ parametrelerinin tahminleri olmak üzere;

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip} \quad i=1,2,\dots,n \text{ olmak üzere}$$

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \text{ minimum olmalıdır.}$$

Çoklu doğrusal regresyon denklemi gözlemler cinsinden aşağıdaki gibi ifade edilir;

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ 1 & x_{21} & x_{22} & \dots & x_{2n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \cdot \\ \cdot \\ \hat{\beta}_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{bmatrix}$$

Y ; $n \times 1$ boyutlu bağımlı değişken vektörü

x ; $n \times (p+1)$ boyutlu bağımsız değişkenlerden oluşan matris

$\hat{\beta}$; $(p+1) \times 1$ boyutlu bilinmeyen parametrelerin tahmini

e ; $n \times 1$ boyutlu hata vektörü olmak üzere;

$\hat{\beta}$ parametrelerinin tahmini değerleri varsayımları sağladığı zaman en küçük kareler yöntemi ile $\hat{\beta} = (X'X)^{-1} X'Y$ şeklinde bulunabilir[9].

2.1.2. En Küçük Kareler Metodunun Varsayımları

Regresyon katsayılarının tahmininde en çok kullanılan En Küçük Kareler Metodunda çeşitli varsayımlar bulunmaktadır. Varsayımlar sağlanmadığı takdirde En Küçük Kareler Metoduyla elde edilen sonuçlar yanlış, tutarsız ve etkisiz olmaktadır[12].

Bu varsayımlar:

1- Hata terimlerinin aritmetik ortalaması sıfır olmalıdır.

$E(e_i)=0$ dır.

Bu varsayımın sağlanmaması durumunda regresyon modeliyle yapılan parametre tahminleri gerçek değerinden, e_i 'lerin pozitif olması durumunda daha büyük, negatif olması durumunda daha küçük olurlar. Yani parametre kestirimleri sapmalı kestirimler olarak elde edilir[20].

2- Hatalar birbirinden bağımsızdır.

$$\text{Cov}(e_i, e_j) = 0$$

Bu varsayım bozulduğu takdirde otokorelasyon problemi ortaya çıkmaktadır. Bu sorunun regresyon analizine çeşitli etkileri bulunmaktadır. Özellikle aralık tahmini ve istatistik testler bağımsızlık ve tesadüfilik varsayımına dayandıkları için geçerliliğini kaybederler, regresyon denkleminin standart hatası ve regresyon katsayılarının standart hataları olması gerekenden düşük çıkabilir[7,20].

3- Hataların varyansı sabittir.

$$\text{Var}(e_i) = \sigma^2$$

Bu varsayım sayesinde X'in bütün değerleri için hata terimleri kendi ortalamaları etrafında aynı dağılımı gösterirler. Bu ise doğrusal regresyon modelinde kestirimlerin standart hatalarının küçük olmasını dolayısıyla daha isabetli olmasını sağlar. Eğer bu varsayım sağlanmazsa değişken varyanslılık sorunu ortaya çıkar. Bu durumda elde edilen regresyon katsayıları yansız olmasına rağmen büyük bir standart hataya sahip olacaktır. Bu ise parametrelere ilişkin güven aralıklarının genişlemesine ve katsayılara ilişkin testlerin düşük duyarlılıkta olmasına neden olacaktır[2,12].

4- Bağımlı değişken ve bağımsız değişkenler arasındaki ilişki doğrusaldır[2]. (model parametreleri için geçerlidir.)

5- Bağımsız değişkenlerin varyansı sıfırdan büyük olmalıdır[2].

$$\text{Var} (X_i) > 0$$

6- Bütün ölçümlerin hatasız yapıldığı varsayılmaktadır[2].

7- n gözlem sayısı, p regresyon modelindeki parametre sayısı olmak üzere $n > p$ olmalıdır.

8- Bağımsız değişkenler arasında anlamlı bir ilişki yoktur[20].

$$\text{Cov} (X_i, X_j) = 0$$

Bağımsız değişkenler arasındaki ilişkinin artması parametre tahminlerinin standart hatalarını yükseltir. Benzer şekilde, değişkenlerin katsayılarının işaretlerinde de farklılaşmaya sebep olabilir. Sonuç olarak bağımlı ve bağımsız değişkenler arasındaki ilişkinin yorumlanmasında yanlış hükümlere varılabilir[14].

En küçük kareler metodu varsayımları sağladığında güvenilir tahminler elde edilmesini sağlamaktadır. İstatistiksel çözümlerinde En Küçük Kareler metodu matematiksel işlemlere en uygun tahmin metodu olarak kullanılsa da varsayımların ihlaline karşı dayanıksızlığı nedeniyle eleştirilmekte ve alternatif olarak daha güçlü yöntemler önerilmektedir[2,20].

2.2. Çoklu Doğrusal Bağlantı

Çoklu regresyon analizinde sıkça karşılaşılan bir sorundur. Bağımsızlık varsayımının bozularak, bağımsız değişkenler arasında doğrusal ilişkinin mevcut olması durumu olarak da özetlenebilir.

Çoklu bağlantı; X ; $n \times (p+1)$ boyutlu veri matrisini göstermek üzere, X_1, X_2, \dots, X_k kolonlarının doğrusal bağımsızlığı olarak da tanımlanabilir. O halde lineer bağımlılık tanımı gereğince t_1, t_2, \dots, t_n en az biri sıfırdan farklı olan skalerler olmak üzere;

$$\sum_{i=1}^p X_i t_i = 0 \text{ ise } X_1, X_2, \dots, X_k \text{ vektörleri lineer bağımlıdır ve tam çoklu bağlantıdan}$$

söz konusudur. Bu durumda $X'X$ matrisinin rankı $(p+1)$ ' den küçük olur ve $(X'X)^{-1}$ hesaplanamaz.

$\sum_{i=1}^p X_i t_i \cong 0$ durumunda ise $(X'X)^{-1}$ hesaplanabilir fakat güçlü bağlantı söz konusu

olduğu için regresyon sonuçları üzerinde olumsuz etkileri olur[21].

Kısaca bağımsız değişkenler arasında tam doğrusal ilişki olduğunda tam çoklu bağlantıdan, bağımsız değişkenler arasındaki bağımsızlığın tam olmadığı durumda kuvvetli çoklu bağlantıdan bahsedilir[11].

Çoklu bağlantı, bir bağımsız değişkenin diğer bağımsız değişkenle olan ilişkisinin derecesine göre tahmin gücünü azaltır. Çoklu bağlantı arttıkça bağımsız değişken tarafından açıklanan spesifik varyans azalmakta, ortak varyans yüzdesi ise artmaktadır. Ortak varyans bir kez sayıldığından, modele yüksek çoklu bağlantılı değişkenler alındıkça genel tahmin gücü daha az artmaktadır[2].

Çoklu bağlantının olduğu durumda en küçük kareler tekniğinin bağımsızlık varsayımı bozulduğu için en küçük kareler yöntemiyle çözüm yapmak uygun olmaz. Çoklu bağlantı varlığında en küçük kareler tahminleri yine yansız fakat çok büyük varyanslı tahminler elde edilir. Böyle durumlarda, regresyon modeliyle yapılacak çıkarsamalar yanlış yönlendirmelere ve hatalara neden olur. Bu durumda çözüm yapabilmek için öncelikle çoklu bağlantının saptanarak hangi çözüm yolunun izleneceğine karar vermek gerekir.

2.2.1. Çoklu Doğrusal Bağlantının Kaynakları

Çoklu doğrusal bağlantı problemi birçok farklı nedenden kaynaklanmaktadır[13]. Bunlar ;

- 1- Örneklem teknikleri: Anakütleyi temsil etme özelliği olmayan örneklemin seçilmesi çoklu bağlantıya sebep olabilir.
- 2- Model ya da anakütle üzerindeki kısıtlamalar: Kullanılan veri toplama tekniğinden bağımsız olarak model ya da anakütleyle ilişkin kısıtlamalar çoklu bağlantıya sebep olabilir. Örneğin elektrik tüketimi ile ilgili olarak, hane gelirinin ve yaşanılan evin büyüklüğünün elektrik tüketimi üzerine etkisi araştırıldığında, hane geliri ile ev büyüklüğü bağımsız değişkenleri arasında çoklu doğrusal bağlantı olduğu görülür. Çoklu doğrusal bağlantının nedeni hane geliri yüksek olan ailelerin daha geniş evlerde oturuyor olduğu gerçeğidir. Burada çoklu doğrusal bağlantının kaynağı anakütlerde var olan gerçek ilişkinin örnekleme de korunuyor olmasıdır.

- 3- Model kurma: İki ya da daha fazla bağımsız değişkenin tama yakın çoklu doğrusal bağlantıya sahip olması durumlarda karşılaşılr. Burada bağımsız değişkenlerin tamamının modelde yer alması çoklu doğrusal bağlantıyı güçlendireceğinden değişkenlerin bir alt kümesi ile çalışılması önerilmektedir.
- 4- Aşırı tanımlanmış model: Bağımsız değişken sayısının gözlem sayısından büyük olduğu modeller çoklu doğrusal bağlantıya sebep olur. Bu durumda; bazı değişkenler modelden dışlanabilir, orijinal bağımsız değişkenlerin sadece alt kümelerini kullanarak ön denemeler yapmak, temel bileşenler yöntemi ile modelden dışlanacak değişkenlerin seçimini yapmak gibi alternatifler denenebilir.

2.2.2. Çoklu Doğrusal Bağlantının Saptanması

Çoklu bağlantının saptanmasında kullanılan çeşitli yöntemler vardır. Bu yöntemler kısaca; Açıklayıcı değişkenler arasındaki korelasyon 1'e yakın olması, $X'X$ matrisin rankı değişken sayısından küçük olması, $X'X$ matrisinin özdeğerleri bir ya da birden fazlası sıfır ya da sıfıra yakın çıkması, Özdeğerlerin terslerinin toplamı bağımsız değişken sayısından çok küçük olması, En büyük özdeğerin en küçük özdeğere bölümü 100'den büyük olması, Standartlaştırılmış $X'X$ matrisinin determinanı sıfır veya sıfıra çok yakın çıkması, j'inci bağımsız değişkenin çoklu belirtme katsayısı R_j^2 değeri 1'e yakın olması, X_j j=1,2,...,k bağımsız değişkenine ait VIF değerleri 10'dan büyük olması, Katsayılarla ilişkin t istatistiklerinin tümü anlamsız iken F istatistiğinin anlamlı çıkması gibi durumlar çoklu bağlantıyı saptamada kullanılan yöntemlerdendir.[11,20,21].

2.2.3. Çoklu Bağlantı Probleminin Sonuçları

- 1- Tam çoklu doğrusal bağlantı durumunda regresyon katsayıları ve bu katsayıların standart hataları sonsuz olmaktadır.
- 2- Güçlü doğrusal bağlantı halinde regresyon katsayıları belirsiz ve bu katsayıların standart hataları sonsuz olmaktadır.
- 3- Modelin R^2 değeri yüksek ancak bağımsız değişkenlerden çok azı t testine göre anlamlı çıkmaktadır.
- 4- Parametre tahminlerinin değeri teorik olarak beklenenden farklı çıkar.

5- Güven aralıkları büyür[2,13,20].

2.2.4. Çoklu Bağlantı Probleminin Çözümü

Çoklu bağlantı probleminin çözümü için; bir veya daha çok bağımsız değişken modelden çıkarılabilir. Fakat hangi değişkenler çıkarılmalıdır? Böyle bir yaklaşım modeli yanlış tanımlamaya götürebilir. Birbiriyle ilişkili olan iki değişken yerine bu iki değişkenin toplamı alınabilir. Bazen yeni gözlem değerlerinin elde edilmesiyle çoklu doğrusal bağlantı problemi ortadan kaldırılabilir. Ancak her zaman örneği büyütme mümkün olmayabilir. Farklar alınarak değişkenler dönüştürülebilir. Fakat böyle bir dönüşüm hatalar arasında otokorelasyon problemine yol açabilir. Ayrıca böyle bir dönüşüm zaman serilerine uygulanabilirken kesit verilerde uygulanamamaktadır. Son olarak, Yanlı tahmin teknikleri kullanılabilir[2].

Çoklu bağlantı sorununu çözmek için kullanılan en etkin yol modeldeki değişkenleri çıkarmadan regresyon katsayılarını yanlı tahmin etmektir. Yanlı tahmin veren yöntemlerden en çok tercih edilenlerin başında, orijinal değişkenler yerine bunların dik dönüşümlerinin kullanıldığı temel bileşenler regresyonu ile korelasyon matrisinin köşegen elemanlarına küçük bir sayının eklenerek kestirim varyanslarının küçültüldüğü Ridge Regresyon gelir. Kısmi en küçük kareler regresyonu da çoklu bağlantı problemiyle temel bileşenler regresyonuna benzer bir yolla başa çıkmaktadır. Bu yanlı tahminler veren yöntemlerin tümü yan miktarında artışa neden olurken, varyansta azalmaya sebep olmaktadır. Bu çalışmada yanlı tahmin yöntemlerinden Temel Bileşenler Regresyonu ve Kısmi en küçük kareler regresyonu incelenecektir.

2.3. Temel Bileşenler Analizi

2.3.1. Temel Bileşenlerin Elde Edilmesi

Çoklu bağlantıyı gidermek için kullanılan yanlı tahmin yöntemlerinden birisi de temel bileşenler regresyonudur. İlk olarak Hotelling (1933) tarafından ele alınmıştır. Bu metotta, korelasyon matrisinin temel bileşenleri olarak adlandırılan yapay değişkenlerin bir kümesi üzerine en küçük kareler yöntemi uygulanır.

Bu çalışmada ilk önce temel bileşenler regresyonunun temelini oluşturan temel bileşenlerin nasıl elde edileceğinden daha sonra ise elde edilen temel bileşenler üzerine en küçük kareler yöntemi uygulanma esasına dayanan temel bileşenler regresyonuna değinilecektir.

Temel bileşenlerin elde edilmesinde, $X_{p \times n}$ ham veri matrisi doğrudan kullanılabilirdiği gibi $Z_{p \times n}$ biçiminde ifade edilen standartlaştırılmış veri matrisi de kullanılmaktadır. Ham veri matrisinin kullanılması durumunda temel bileşenlerin bulunmasında varyans-kovaryans matrisinden, standartlaştırılmış veri matrisinin kullanılması durumunda korelasyon matrisinden yararlanılmaktadır. Farklı sonuçlar verebilen bu iki yoldan hangisinin seçileceği konusunda en önemli belirleyici, değişkenlerin ölçü birimleridir. Eğer değişkenlerin ölçü birimleri ve varyansları birbirine yakınsa kovaryans matrisinden, değilse korelasyon matrisinden yararlanır. Değişkenlerin ölçü birimlerinin birbirine yakın olması pratikte pek rastlanmaz.

Y, bağımlı değişkeni, X bağımsız değişkenleri, β , regresyon katsayılarını ve e hata terimini gösteriyor olsun:

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ X_{p1} & X_{p2} & \dots & X_{pn} \end{bmatrix}$$

X, (pn) boyutunda veri matrisidir. Bu matriste, n; toplam gözlem sayısını, p; toplam değişken sayısını, göstermektedir.

Temel bileşenler regresyonunda ilk olarak hem bağımlı hem de bağımsız değişkenler ortalamadan farkları alınıp standart sapmalarına bölünerek standartlaştırılmaktadır.

$$i^{\text{inci}} \text{değişkenin aritmetik ortalaması: } \bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}$$

$$i\text{'inci deęişkenin standart sapması: } S_i = \sqrt{\sum_{j=1}^n \frac{(X_{ij} - \bar{X}_i)^2}{n}}$$

$$i \text{ deęişkeninin } j \text{ gözlemi için standartlaştırılmış deęeri: } Z_{ij} = \frac{(X_{ij} - \bar{X}_i)}{S_i}$$

$i = 1, 2, \dots, p \quad j = 1, 2, \dots, n$ olmak üzere:

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{1n} \\ Z_{21} & Z_{22} & \dots & Z_{2n} \\ \dots & \dots & & \dots \\ \dots & \dots & & \dots \\ \dots & \dots & & \dots \\ Z_{p1} & Z_{p2} & \dots & Z_{pn} \end{bmatrix}$$

$Z = (Z_1, Z_2, \dots, Z_p)$ vektörlerinden oluşan $(p \times n)$ boyutta standartlaştırılmış veri matrisi elde edilir. Bu şekilde elde edilen standartlaştırılmış deęişkenlerin aritmetik ortalaması 0, standart sapması 1 dir.

Z_i ile Z_k arasındaki kovaryans:

$$Cov(Z_i, Z_k) = S_{ik} = \frac{1}{n-1} \sum_{j=1}^n ((Z_{ij} - \bar{Z}_i)(Z_{kj} - \bar{Z}_k)) \text{ olduğundan,}$$

i deęişkeni ile k deęişkeni arasındaki korelasyon katsayısı;

$$r_{ik} = \frac{Cov(Z_i, Z_k)}{\sqrt{Var(Z_i)}\sqrt{Var(Z_k)}} = \frac{\sum_{j=1}^n ((Z_{ij} - \bar{Z}_i)(Z_{kj} - \bar{Z}_k))}{\sqrt{\sum_{j=1}^n (Z_{ij} - \bar{Z}_i)^2} \sqrt{\sum_{j=1}^n (Z_{kj} - \bar{Z}_k)^2}} \quad \begin{array}{l} i = 1, 2, \dots, p \\ k = 1, 2, \dots, p \\ r_{ik} = r_{ki} \\ i \neq k \end{array}$$

$$R = \frac{ZZ'}{n-1}$$

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \dots & \dots & & \dots \\ \dots & \dots & & \dots \\ \dots & \dots & & \dots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix}$$

R, (p x p) boyutunda bağımsız değişkenler için korelasyon matrisini göstermektedir.

$\lambda_1, \lambda_2, \dots, \lambda_p$ 'ler, $\det(R - \Lambda I) = 0$ eşitliğini sağlayan, ve $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ koşuluna uyan özdeğerlerdir.

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \lambda_2 & \dots & \dots & \dots & 0 \\ \dots & \dots & & & & \dots \\ \dots & \dots & & & & \dots \\ \dots & \dots & & & & \dots \\ 0 & 0 & & & & \lambda_p \end{bmatrix}$$

$a = (a_1, a_2, \dots, a_p)$ vektörleri ise sıfırdan farklı, $Ra_i = \lambda_i a_i$ eşitliğinden $a_i' a_i = 1$ ve $a_i' a_j = 0$ ($i \neq j = 1, \dots, p$) şartlarının sağlanmasıyla elde edilen, korelasyon matrisinin standartlaştırılmış özvektörleridir.

$$a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \dots \\ \dots \\ \dots \\ a_{p1} \end{bmatrix} \quad a_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \dots \\ \dots \\ \dots \\ a_{p2} \end{bmatrix} \quad \dots \quad a_p = \begin{bmatrix} a_{1p} \\ a_{2p} \\ \dots \\ \dots \\ \dots \\ a_{pp} \end{bmatrix}$$

a' ise a vektörünün devrik (transpoze) vektörüdür.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \dots & \dots & & \dots \\ \dots & \dots & & \dots \\ \dots & \dots & & \dots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{bmatrix}$$

Temel bileşenler regresyonu, birbiriyle ilişkili Z_{ij} değerlerinin bir dönüşüm yardımıyla birbiriyle ilişkisiz W_{ij} değerlerine ulaşmamızı sağlamaktadır.

$$W_{p \times n} = A'_{p \times p} Z_{p \times n}$$

Genel olarak bu şekilde gösterilebilir.

$$W_1 = (a_1)'Z = a_{11} Z_1 + a_{21} Z_2 + \dots + a_{p1} Z_p$$

$$W_2 = (a_2)'Z = a_{12} Z_1 + a_{22} Z_2 + \dots + a_{p2} Z_p$$

...

...

$$W_p = (a_p)'Z = a_{1p} Z_1 + a_{2p} Z_2 + \dots + a_{pp} Z_p$$

a_{ij} 'ler her bir temel bileşenin hangi değişkenle hangi oranda ilişkilendirildiğini gösteren özvektörlerdir. Temel bileşen yükleri olarak da adlandırılırlar. Temel bileşenlerin varyansları ve kovaryansları;

$$\text{Var}(W_i) = \text{Var}((a_i)'Z) = (a_i)'R a_i$$

$$\text{Cov}(W_i, W_k) = (a_i)'R a_k$$

Burada $W_1 = (a_1)'Z$ dönüştürülmüş vektöre birinci temel bileşen adı verilir. a_1 vektörüne ise birinci özvektör adı verilir.

W_1, W_2, \dots, W_p Temel bileşenleri belirlenirken birbirinden bağımsız olarak;

$\text{Var}(W_1) > \text{Var}(W_2) > \dots > \text{Var}(W_p)$ olacak şekilde seçilmelidir.

Bu durumda toplam varyansa katkısı en fazla olan birinci temel bileşen;

$W_1 = (a_1)'Z = a_{11} Z_1 + a_{21} Z_2 + \dots + a_{p1} Z_p$ doğrusal birleşimidir.

$\text{MaxVar}(W_1) = (a_1)' R a_1$ eşitliğinde a_1 vektörü birinci temel bileşenin varyansını maksimum yapacak şekilde belirlenmektedir. Ancak a_1 vektörü herhangi bir sabit sayı ile çarpılarak, değişkenlik hiçbir kısıtlamaya bağlı kalmaksızın arttırılabilir. Bundan dolayı a_i vektörleri birim uzunlukta $(a_i)' \cdot a_i = 1$ seçilmesi uygun olacaktır. Bu şekilde seçilen;

- 1- Birinci bileşen $\text{MaxVar}(W_1) = (a_1)' Z$ ve $(a_1)' \cdot a_1 = 1$ şartını sağlayan $(a_1)' \cdot Z$ doğrusal bileşendir.
- 2- İkinci bileşen $\text{max Var}((a_2)' Z)$ ile $(a_2)' \cdot a_2 = 1$ ve $\text{Cov}(W_1, W_2) = 0$ şartını sağlayan $(a_2)' \cdot Z$ doğrusal bileşendir.
- 3- i'inci temel bileşen $\text{max Var}((a_i)' Z)$ ile $(a_i)' \cdot a_i = 1$ ve $i > k$ olmak üzere $\text{Cov}(W_i, W_k) = 0$ şartlarını sağlayan $(a_i)' Z$ doğrusal bileşimidir.

$W_i = a_i' Z$ $i=1,2,\dots,p$ olmak üzere;

$$\sum_{i=1}^p \text{Var}(W_i) = \sum_{i=1}^p \text{Var}(Z_i) = p$$

$$\rho_{Y_i, Z_k} = a_{ki} \sqrt{\lambda_i} \quad i=1,2,\dots,p \quad k=1,2,\dots,p$$

Standartlaştırılmış değişkenlerin toplam kitle varyansı p 'dir. Aynı zamanda standartlaştırılmış veri matrisinin korelasyon matrisinin (R) determinanı, p tane özdeğerin

toplamına, yani p 'ye eşittir ($|R| = \prod_{i=1}^p \lambda_i$).

Birinci temel bileşenin toplam varyansı açıklama oranı;

$$\frac{\lambda_1}{p} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$$

Birinci ve ikinci temel bileşenin toplam varyansı açıklama oranı;

$$\frac{\lambda_1 + \lambda_2}{p} = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$$

p tane temel bileşenin toplam varyansı açıklama oranı;

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{p} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{\lambda_1 + \lambda_2 + \dots + \lambda_p} = 1 \text{ dir.}$$

Kısaca temel bileşenler analizinde, veri matrisi standartlaştırılarak korelasyon matrisi bulunur. Korelasyon matrisinin özdeğerleri ve standartlaştırılmış özvektörleri bulunur. Her

bir özvektörün devrik vektörü elde edilir. Devrik özvektör standartlaştırılmış veri matrisi ile çarpılarak temel bileşen değerleri bulunur [6,10,17,18].

2.3.2. Temel Bileşenlerin Özellikleri

1- $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ dir

2- $|R| = |A\Lambda A'| = |A||\Lambda A'| = |A||\Lambda||A'| = |\Lambda||A||A'| = |\Lambda| = \prod_{j=1}^p \lambda_j$ burada $\Lambda: p \times p$

boyutlu köşegen elemanları λ_j 'ler, köşegen dışı elemanları 0 olan bir matristir.

$\text{izR} = \sum_{j=1}^p \Lambda = p$ olduğunu açıklar.

3- Herhangi bir bileşenin toplam varyansı açıklama oranı $(\lambda_i / \sum_{i=1}^p \lambda_i) \times 100$ dür.

4- $W_i = a_i' Z$ ($i=1,2,\dots,p$) eşitliğini kullanarak

$$\text{var}(W_i) = \frac{W_i W_i'}{n-1} = (a_i' Z)(a_i' Z)' / (n-1) = a_i' [ZZ' / (n-1)] a_i = a_i' R a_i =$$

$a_i' \lambda_i a_i = \lambda_i a_i' a_i = \lambda_i$ buradan özdeğerlerin temel bileşenlerin varyansları olduğunu söyleyebiliriz.

5- Orijinal değişkenlerin toplam varyansı temel bileşenlerin toplam varyansına eşittir.

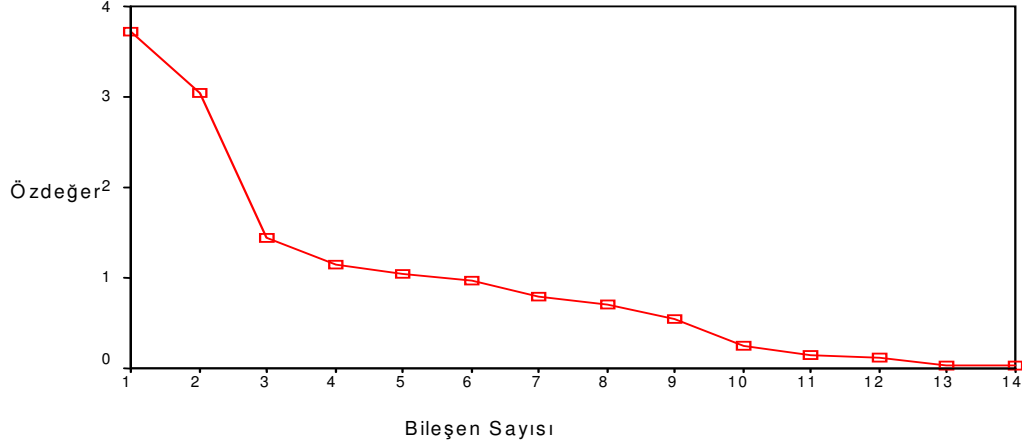
$$\sum_{i=1}^p \text{var}(z_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{var}(y_i)$$

şeklindedir [21].

2.3.3. Temel Bileşen Sayısının Belirlenmesi

R veri matrisinin özdeğerlerinin bulunmasından sonra q önemli özdeğer sayısına karar vermek çok önemlidir. q sayısının belirlenmesinde birkaç yaklaşım bulunmaktadır. Bunlardan en basiti standartlaştırılmış veri matrisinin kullanıldığı durumlarda 1 den büyük özdeğerlerin sayısını almaktır. Ya da başka bir ifade ile $\sum_{j=1}^q \lambda_j / p \geq 2/3$ koşulunun sağladığı en küçük q değeri önemli bileşenlerin sayısı olarak alınabilmektedir.

Temel bileşen sayısını belirlemede kullanılan bir diğer yöntem ise grafik yöntemidir. Özdeğerlerin çizimi yoluyla(scree graph), varyans açıklama oranlarındaki hızlı düşüş belirlenerek, bileşen sayısına karar verilebilir.



Şekil 1. Özdeğerlerin Varyans Açıklama Oranları

Grafik yöntemine göre karar şöyle verilir; en hızlı düşüş ve toplam varyansı açıklama oranını kriter olarak kullanılır. İlk 5 değişken anlamlı bulunur[17].

2.4. Temel Bileşenler Regresyonu

Çoklu bağlantıyı gidermek için kullanılan temel bileşenler regresyonunda, orijinal değişkenler yerine bunların dik dönüşümü kullanılarak elde edilen yapay değişkenlerin bir kümesi üzerine, en küçük kareler yöntemi uygulanarak regresyon katsayılarının ($\hat{\beta}$) tahmini yapılır. Temel bileşenler regresyonuyla yapılacak tahminde hata kareleri ortalamasının en küçük kareler yöntemiyle yapılacak tahmine göre daha küçük olması beklenmektedir.

Matris notasyonu ile verilen regresyon modeli dikkate alınır;

$$Y = X\beta + \varepsilon$$

En Küçük Kareler Yönteminde $\hat{\beta}$ regresyon katsayılarının tahmini;

$$\hat{\beta} = (X'X)^{-1} X'Y \quad \text{şeklindedir.}$$

Bağımsız değişkenler, temel bileşenlere aşağıdaki gibi dönüştürülür.

$$X'X = PDP' = W'W$$

Burada;

D; X'X özdeğerlerinin köşegen matrisini,

P; X'X özvektör matrisini göstermektedir.

Temel bileşenler seti üzerine uygulanan en küçük kareler tahmini;

$$\gamma = (W'W)^{-1}W'Y \quad \text{şeklindedir. Böylece modelde çoklu doğrusal bağlantı}$$

olmamaktadır.

γ temel bileşenler regresyonuyla tahmin edilen regresyon katsayısıdır.

γ ve $\hat{\beta}$ regresyon katsayıları seti arasındaki ilişki;

$$\gamma = P' \hat{\beta} \quad \text{ve} \quad \hat{\beta} = P\gamma \quad \text{şeklindedir.}$$

$\hat{\beta} = P\gamma$ eşitliği kullanılarak hesaplanan regresyon katsayıları orijinal ölçeğine geri çevrilir[10,17,18].

2.4.1. Teorik Model

$Y_i = \beta_0 + X\beta$ regresyon modeli dikkate alınır, elde edilen temel bileşenlerle model yeniden oluşturulursa;

$$\begin{aligned} Y_i &= \beta_0 + X\beta = \beta_0 + \sum_{j=1}^p X_{ij}\beta_j = \gamma_0 + W_{ik}(\gamma_k) = \gamma_0 + \sum_{k=1}^p W_{ik}\gamma_k \\ &= \gamma_0 + \sum_{k=1}^p \sum_{j=1}^p \gamma_k \left(\frac{X_{ij}}{S_j} - \frac{\bar{X}_j}{S_j} \right) a_{jk} = \gamma_0 + \sum_{k=1}^p \sum_{j=1}^p \frac{\gamma_k a_{jk}}{S_j} X_{ij} - \sum_{k=1}^p \sum_{j=1}^p \frac{\gamma_k a_{jk}}{S_j} \bar{X}_j \\ &= \sum_{j=1}^p X_{ij} \left(\frac{\gamma_k a_{jk}}{S_j} \right) + \gamma_0 - \sum_{k=1}^p \gamma_k \sum_{j=1}^p \frac{a_{jk}}{S_j} \bar{X}_j \quad \text{olduğu görülür.} \end{aligned}$$

Regresyon katsayıları arasındaki ilişki;

$$\beta_j = \sum_{k=1}^p \left(\frac{\gamma_k a_{jk}}{S_j} \right) \quad \text{ve} \quad \beta_0 = \gamma_0 - \sum_{k=1}^p \gamma_k \sum_{j=1}^p \frac{a_{jk}}{S_j} \bar{X}_j \quad \text{olduğu açıktır[4].}$$

2.5. Kısmi En Küçük Kareler Regresyonu

Kısmi en küçük kareler regresyonu, Kısmi en küçük kareler analizi ve çoklu doğrusal regresyon analizinden oluşan çok değişkenli istatistiksel bir yöntemdir. Kısmi en küçük kareler analizi ile fazla sayıda olan ve aralarında çoklu doğrusal bağlantı bulunan açıklayıcı değişkenler, açıklayıcı ve yanıt değişkenindeki değişimi büyük ölçüde açıklayan az sayıda ve aralarında çoklu doğrusal bağlantı bulunmayan yeni değişkenlere indirgenmektedir. Elde edilen bileşenlere çoklu regresyon analizi uygulanarak regresyon modeli oluşturulmaktadır.

Kısmi en küçük kareler regresyonunda, temel bileşenler regresyonundan farklı olarak hem açıklayıcı hem de yanıt değişkenindeki bilgiden yararlanmaktadır. Temel bileşenler regresyonundan daha az bileşenli modeller kurmayı amaçlamaktadır. Bu amaçla kullanılan kısmi en küçük kareler regresyonunda çeşitli algoritmalar bulunmaktadır. Çalışmada uygulama NIPALS algoritmasıyla yapıldığı için sadece bu algoritmaya değinilecektir[3,5,21].

2.5.1. NIPALS Algoritması

Klasik algoritma olarak da bilinen NIPALS tek bağımlı değişken ve çok bağımlı değişken olduğu durumlarda kullanılabilir. Kovaryans matrisini en çoklayan bileşenleri elde etmeyi amaçlayan algoritmada tüm bileşenler aynı anda elde edilemez. Her bir adımda tek bir bileşen ve bu bileşenlere ait ağırlık ve yük değerleri elde edilmektedir. Algoritma istenilen bileşen sayısı elde edilinceye ya da X matrisi sıfır matris olunca sonlandırılır. NIPALS algoritması , $N \times K$ boyutlu X açıklayıcı değişkenler matrisi ve $N \times P$ Boyutlu Y yanıt değişkenler matrisi ile ilgilenmektedir. Burada K: açıklayıcı değişken sayısını, P: bağımlı değişken sayısını vermektedir. Algoritmada a bileşen sayısını göstermekte olup $a = 1, 2, \dots, A$ dır. İlk adımda orijinal matrislerin kullanıldığı algoritma aşağıdaki adımlardan oluşmaktadır[3,5,23].

1. Bağımlı değişken çok sayıda ise bu değişkenlerden oluşan Y matrisinin en yüksek varyansa sahip olan sütunu ya da ilk sütunu, Bağımlı değişken tek ise o değişken sütunu u_a vektörü olarak alınır. $u_a = Y$ dir.

2. X ve Y'nin ilgili bileşeni u_a üzerine regresyonundan X ve u arasındaki kovaryansı en çoklayan w ağırlık vektörü $w_a = X_a' u / (u_a' u_a)$ ile elde edilir.
3. $w_a / \|w_a\|$ ile w_a vektörü normuna bölünerek boyu 1 olacak şekilde ölçeklendirilir.
4. X skorları olan t_a , $t_a = X_a w_a$ eşitliği ile hesaplanır.
5. t_a bileşeninin Y'yi modellemedeki katsayısını açıklayan c_a ağırlık vektörü; $c_a = Y_a' t_a / (t_a' t_a)$ ile Y'nin t_a üzerine regresyonundan elde edilir.
6. c_a ağırlık vektörü normuna bölünerek boyu 1 olacak şekilde ölçeklendirilir. Yani $c_a / \|c_a\|$ hesaplanır.
7. Y skorlarının güncellenmiş bir kümesi $u_{a(yeni)}$, c_a ağırlık vektörü ile Y'nin doğrusal bir kombinasyonu olacak şekilde $Y_a c_a / (c_a' c_a)$ ile hesaplanır.
8. Adım 2'de kullanılan u_a değeri ile Adım 7'de kullanılan $u_{a(yeni)}$ arasında bir yakınsama sağlanıp sağlanmadığına bakılır. Bu yakınsama, iki vektörün farkının normunun 10^{-6} gibi sifıra çok yakın bir değer olması ile tespit edilir. Bu yakınsama sağlanır ise sonraki adımlara geçilerek algoritma sonlandırılır. Sağlanmazsa Adım 7'de elde edilen $u_{a(yeni)}$ değeri Adım 2'de yerine koyularak algoritmaya devam edilir. Bir tek Y değişkeni varsa süreç tek bir yinelemede yakınsar ve Adım 9'dan devam eder.
9. X'in ilgili bileşeni t_a üzerine regresyonundan, bileşenin açıklayıcı değişken üzerindeki etkisini ifade eden yük vektörü p_a , $p_a = X_a' / (t_a' t_a)$ ile elde edilir.
10. Y'nin ilgili bileşeni u_a üzerine regresyonundan, bileşenin bağımlı değişken üzerindeki etkisini ifade eden yük vektörü q_a , $q_a = Y_a' u_a / (u_a' u_a)$ ile elde edilir.
11. hem X hem de Y için bileşenler ayrı hesaplandığından bileşenler arasında zayıf bir ilişki olmaktadır. Bu durumu ortadan kaldırmak için her bir bileşen için Y'nin ilgili bileşeni u_a 'nın X'in ilgili bileşeni t_a üzerine regresyonundan elde edilen içsel bir ilişkiyi tanımlayan $b_a = u_a' t_a / (t_a' t_a)$ ile hesaplanır.
12. Elde edilen bileşenler ve yükler bağımlı ve açıklayıcı değişkeni modellemede kullanılmaktadır. Sırasıyla açıklayıcı ve bağımlı değişken $X = TP'$ ve $Y = BTC'$ ile modellenmektedir. Algoritmanın bu adımda bir sonraki bileşeni

elde etmek için kullanılacak olan X_{a+1} ve Y_{a+1} artık matrisleri $X_{a+1} \rightarrow X_a - t_a p_a'$ ve $Y_{a+1} \rightarrow Y_a - b t_a c_a'$ ile hesaplanmaktadır. Algoritma açıklayıcı değişkenlerdeki ve bağımlı değişkenlerdeki değişimin büyük bir kısmı açıklanmaya kadar devam edilir. Algoritma ihtiyaç duyulan en az sayıda bileşen sayısını vermektedir.

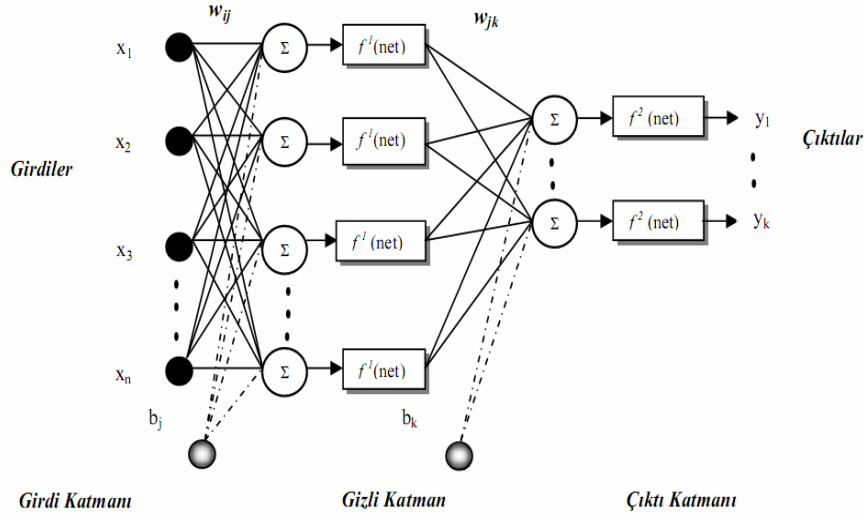
2.6. Yapay Sinir Ağları

Yapay sinir ağları, insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacı ile geliştirilen bilgisayar sistemleridir. Bu yetenekleri geleneksel programlama yöntemleri ile gerçekleştirmek oldukça zor veya mümkün değildir. Bu nedenle yapay sinir ağlarının, programlanması çok zor veya mümkün olmayan olaylar için geliştirilmiş adaptif bilgi işleme ile ilgilenen bir bilgisayar bilim dalı olduğu söylenebilir.

Yapay sinir ağları; öğrenme, ilişkilendirme, sınıflandırma, özellik belirleme, genelleme, optimizasyon gibi konularda başarılı bir şekilde uygulanabilirler.

2.6.1. Yapay Sinir Ağlarının Yapısı

Genel olarak yapay sinir ağı üç katman halinde tanımlanmaktadır. İlk katman giriş katmanı son katman ise çıkış katmanıdır. Aradaki katmanlar ise gizli katman olarak tanımlanır. Yapay sinir ağında çok sayıda gizli katman bulunabilir. Bir yapay sinir ağında kaç tane gizli katman olacağı ve her bir gizli katmanda kaç sinir hücresi olacağı bugüne kadar belirlenememiştir. Probleme göre değişen bu duruma deneme yanılma yoluyla çözüm getirilebilmiştir.



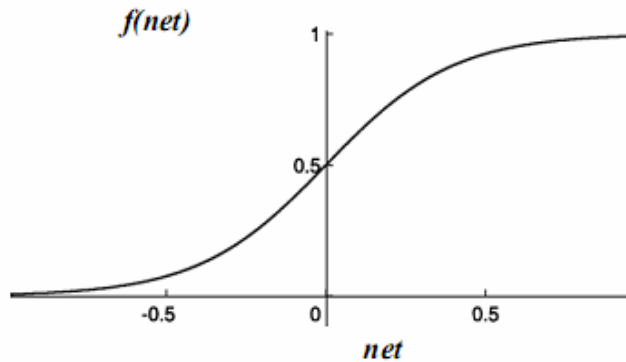
Şekil 2. Yapay Sinir Ağı Yapısı

x_i : $i=1,2,\dots,n$ ağın girdileridir.

w_{ij} : girdi katmanı ile gizli katman arasındaki, w_{jk} : çıktı katmanı ile gizli katman arasındaki ağırlıkları gösterir. Girdinin hücre üzerindeki etkisini göstermeye yarar. Gelen girdi değerleri kendi ağırlığı ile çarpılarak net girdi değeri elde edilir.

b_j : girdi katmanı ile gizli katman arasındaki, b_k : çıktı katmanı ile gizli katman arasındaki eşik değerlerdir. Bilginin iletimi için gerekli olan minimum değeri gösterir.

f_1 : girdi katmanı ile gizli katman arasındaki, f_2 : çıktı katmanı ile gizli katman arasındaki aktivasyon fonksiyonudur. Hücreye gelen net girdiyi belirleyerek hücrenin üreteceği çıktıyı belirlemeye yarar. Günümüzde çok katmanlı algılayıcı modellerde yaygın olarak sigmoid aktivasyon fonksiyonu kullanılmaktadır.



Şekil 3. Sigmoid Fonksiyon Grafiği

Sigmoid fonksiyon kullanıldığında hücre çıktısı $f_{(net)} = \frac{1}{1 + e^{-net}}$ formülüyle

hesaplanır. Burada net; net girdi değerini göstermektedir.

j: gizli katmana ait hücre sayısını gösterir.

n: girdi katmanındaki hücre sayısını gösterir.

k: çıkış katmanındaki hücre sayısını gösterir.

İleri beslemeli ağlarda; nöronlar girişten çıkışa doğru düzenli katmanlar şeklindedir. Bir katmandan sadece kendinden sonraki katmanlarla bağ bulunmaktadır. Bilgiler önce giriş katmanında, sonra ara katmanda ve daha sonra çıkış katmanında işlenerek çıkışlar elde edilir. İleri beslemeli ağda girişleri çıkışlara çeviren fonksiyon aşağıdaki gibidir.

$$y_k = f_2 \left\{ \sum_{j=1}^k [f_1 \left(\sum_{i=1}^n x_i w_{ij} + b_j \right)] w_{jk} + b_k \right\} \text{ formülü ile hesaplanır.}$$

Geri beslemeli ağlarda; ağın ürettiği çıktı değeri beklenen çıktı değeri ile karşılaştırılmaktadır. Çıktı ile beklenen değerler arasındaki fark yapay sinir ağı modellerinde hata olarak adlandırılmakta ve geriye yayılım aşamasında bu hatanın kabul edilebilir düzeye indirilebilmesi ağın ağırlık ve eşik değerlerinin iteratif olarak değiştirilmesi ile mümkün olmaktadır. Böylece başlangıçta rastgele atanan ağın ağırlıkları hata istenen sınırlara ulaşıncaya kadar güncellenmektedir. Bu yapay sinir ağı modellerinde hata kareler toplamı veya ortalaması hesabına dayalı amaç fonksiyonu ile gerçekleştirilebilmektedir.

$$E(w) = \frac{1}{2} \sum_{i=1}^n e_i^2(w) \text{ Burada; } E(w) \text{ amaç fonksiyonu, } e(w) \text{ ise çıktı ile beklenen}$$

değer arasındaki farkı göstermektedir.

Bu çalışmada ileri beslemeli geri yayımlı ağ algoritmasının gelişmiş bir türü olan Levenberg-Marquardt eğitim algoritmasından yararlanılacaktır. Levenberg-Marquardt algoritmasında Hessian matrisi ($H(w)$) adı verilerin çözümlenmesi karışık bir matrisin yaklaşık değeri kullanılmaktadır. Levenberg-Marquardt yönteminin asıl amacı amaç ölçütünün ikinci türevini alarak bulunan H matrisini hesaplamaksızın jakobiyenden yararlanarak yaklaşık H matrisini hesaplatmaktır.

$$H(w) \cong J^T(w) J(w) + \mu I$$

μ : Marquardt parametresi

I : Birim matrisi

J : Jakobien matristir. Ağ hatalarının ağırlıklara göre birinci türevlerinden elde edilmektedir.

Ağın hatasının geri yayılması aşamasında, öncelikle Jakobien matrisin transpozu ve ağ hataları kullanılarak ağın gradyeni hesaplanmaktadır;

$$\nabla E(w) = J^T(w) e(w)$$

Ağın gradyeni hesaplandıktan sonra, ağın ağırlıklarındaki vektörel değişim Hessian matrisinin tersi ile ağın gradyeninin çarpılmasıyla belirlenmekte ve ağın ağırlıkları güncellenmektedir.

$$\Delta w = -[H(w)]^{-1} \nabla E(w)$$

$$w_{\text{yeni}} = w_{\text{eski}} + \nabla w$$

[24,15]

3. UYGULAMA

3.1. Gereç ve Yöntem

3.1.1. Araştırmanın Tasarımı ve Modeli

Bu araştırma hipertansif hastaların sol atriyum çapı üzerinde etkili olabileceği düşünülen değişkenlerin etkisinin değerlendirilmesine yönelik kesitsel bir araştırmadır. Verilerin toplanması ileriye yönelik olarak gerçekleştirilmiştir. Değişkenler arasında çoklu doğrusal bağlantılar aranmıştır. En küçük kareler regresyonu, temel bileşenler regresyonu, kısmi en küçük kareler regresyonu ve yapay sinir ağları ile incelenerek sol atriyum çapı için en iyi tahmin elde edilmeye çalışılmıştır. Çalışmada sol atriyum çapını tahmin etmek için kullanılan değişkenler ve ölçü birimleri aşağıdaki tabloda özetlenmiştir.

Tablo 3.1. Modeldeki Değişkenler ve Ölçü Birimleri

Değişkenin Türü	Değişkenler		Ölçü Birimi
Bağımsız	x_1	Yaş	Yıl
	x_2	İzovolümetrik Relaksasyon Zamanı	ms
	x_3	Protrombin Zamanı	sn
	x_4	Erken Diastolik Akım Hızı	cm/sn
	x_5	Sistolik Enjeksiyon	cm/sn
	x_6	Sistolik Çap	mm
	x_7	Diastolik Çap	mm
	x_8	Sol VAD kalınlık	mm
	x_9	Aort Kökü Çapı	mm
	x_{10}	Sistolik Kan Basıncı	mm Hg
	x_{11}	Diastolik Kan Basıncı	mm Hg
	x_{12}	Enjeksiyon Fraksiyonu	%
	x_{13}	Ortalama Arter Basıncı	mm Hg
Bağımlı	Y	Sol Atriyum Çapı	mm

3.2. Bulgular

Elazığ Fırat Üniversitesi Tıp Fakültesi Hastanesi kardiyoloji polikliğine gelen 127 hastadan elde edilen veriler ekler bölümünde gösterildiği gibidir. Elde edilen sonuçlar tablolar halinde verilmiş ve yorumları tablonun altında özetlenmiştir. Çoklu doğrusal regresyon ve temel bileşenler regresyonu için Ncss 2007 programından faydalanılmıştır.

Tablo 3.2. Tanımlayıcı istatistikler

Değişkenler	n	\bar{x}	s(std. sapma)	Min	Max
x_1	127	52.141	11.345	20	75
x_2	127	82.685	21.937	21	224
x_3	127	46.551	7.278	32	70
x_4	127	0.0456	3.061	0.01	0.17
x_5	127	0.036	2.134	0.01	0.12
x_6	127	28.133	3.708	21	45
x_7	127	44.110	3.210	28	57
x_8	127	9.181	1.157	7	13
x_9	127	32.897	2.869	25	41
x_{10}	127	140.866	13.215	110	170
x_{11}	127	88.425	7.498	70	110
x_{12}	127	56.417	4.313	45	65
x_{13}	127	105.906	8.900	83.333	130
Y	127	36.094	3.200	30	45

Tablo 3.2'de 127 adet veri için her bir değişkenin ortalaması, standart sapması, minimum ve maksimum değerleri görülmektedir.

Tablo 3.3. EKK Çoklu Regresyon Sonuçları

Bağımsız değişkenler	Regresyon Katsayıları	Standart Hata	t-testi	p
Sabit	22.611	7.472	3.026	0.003*
x_1	0.043	0.025	1.680	0.095
x_2	0.024	0.0121	1.98	0.049
x_3	-0.063	0.036	-1.771	0.079
x_4	-2.853	9.739	-0.293	0.770
x_5	-5.472	13.011	-0.421	0.674
x_6	0.162	0.084	1.921	0.057
x_7	0.100	0.094	1.069	0.287
x_8	0.757	0.333	2.273	0.024*
x_9	0.101	0.099	1.016	0.319
x_{10}	41.127	32.576	1.262	0.209
x_{11}	82.284	65.140	1.263	0.209
x_{12}	-0.076	0.066	-1.164	0.246
x_{13}	-123.428	97.714	-1.236	0.209

Tablo.3.3’de EKK analiziyle elde edilen regresyon katsayıları, standart hataları ve regresyon katsayıları için t-testi sonuçları verilmiştir.

Tablo 3.4. EKK için ANOVA tablosu

Değ. Kay.	sd	HKT	HKO	F	p
Sabit	1	165457.100	165457.100	4.933	<0.0001
Model	13	467.341	35.949		
Hata	113	823.524	7.280		
Toplam	126	1290.866	10.244		

Tablo3.4’de modele ilişkin ANOVA varyans analiz tablosu verilmiştir. Tablo3.4’e göre $p < 0.05$ olduğu için model anlamlıdır. Fakat tablo3.3’e bakıldığında t- testi sonuçlarına göre katsayıların büyük çoğunluğu anlamsız çıkmaktadır. F testi anlamlı iken t-testi sonuçlarının anlamsız çıkması çoklu bağlantı probleminin bir sonucudur. Bu nedenle bağımsız değişkenler arasında çoklu bağlantı olup olmadığının incelenmesine karar verilmiştir.

Tablo 3.5. Korelasyon Matrisi

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	Y
x_1	1													
x_2	0.081	1												
x_3	-0.070	0.105	1											
x_4	-0.422	-0.074	0.228	1										
x_5	-0.165	-0.059	0.199	0.373	1									
x_6	0.052	0.136	-0.173	-0.207	-0.013	1								
x_7	-0.249	0.209	-0.115	-0.085	-0.184	0.407	1							
x_8	0.069	-0.003	-0.026	-0.217	-0.213	0.107	0.116	1						
x_9	-0.008	0.133	-0.199	-0.119	-0.093	0.442	0.226	0.268	1					
x_{10}	0.072	0.070	-0.048	-0.259	-0.197	0.075	-0.022	0.731	0.236	1				
x_{11}	0.086	0.052	-0.114	-0.324	-0.211	0.113	0.050	0.563	0.198	0.790	1			
x_{12}	-0.304	-0.159	0.175	0.285	0.227	-0.331	-0.028	-0.020	-0.206	-0.098	-0.040	1		
x_{13}	0.084	0.064	-0.088	-0.310	-0.216	0.100	0.016	0.678	0.228	0.939*	0.953*	-0.071	1	
Y	0.220	0.252	-0.233	-0.284	-0.206	0.370	0.256	0.323	0.318	0.217	0.198	-0.285	0.218	1

Tablo3.5'te her bir bağımsız değişkenin birbirleriyle ve bağımlı değişkenle aralarındaki basit kolerasyon katsayılarından oluşan basit kolerasyon matrisi verilmektedir. Çoklu bağlantıyı belirlemek için kullanılan yöntemlerden biri de korelasyon matrisinin incelenmesidir. Genel olarak ve bağımsız değişkenleri lineer bağımlı olduklarında olur. Tablo3.5'te $x_{10} - x_{13}$ değişkenleri arasındaki kolerasyon 0.939, $x_{11} - x_{13}$ arasındaki kolerasyon 0.953 olduğu için çoklu bağlantı probleminin varlığından söz edilebilir. Fakat basit kolerasyon çoklu bağlantının belirlenmesinde tek başına yeterli değildir. Bu nedenle çoklu bağlantıyı belirleyen diğer durumların da incelenmesine karar verilmiştir.

Tablo 3.6. EKK çoklu bağlantının saptanması

Bağımsız değişkenler	VIF	Diğer değişkenlere olan R^2	Tolerans
x_1	1.484	0.326	0.673
x_2	1.223	0.182	0.817
x_3	1.179	0.152	0.847
x_4	1.611	0.379	0.620
x_5	1.336	0.250	0.749
x_6	1.706	0.413	0.586
x_7	1.585	0.369	0.630
x_8	2.575	0.611	0.388
x_9	1.405	0.288	0.711
x_{10}	3204156.888*	1*	0*
x_{11}	41258972.939*	1*	0*
x_{12}	1.401	0.286	0.713
x_{13}	13078203.449*	1*	0*

Tablo3.6'da bağımsız değişkenlere ait VIF, ve tolerans değerleri verilmiştir. Çoklu bağlantıyı belirleme yöntemlerinden biri olan VIF için VIF >10 olduğu zaman çoklu bağlantı problemi ortaya çıkmaktadır. Tablo3.6'ya göre değişkenlerine ait x_{10} , x_{11} ve x_{13} değişkenlerinin VIF değerlerinin 10 dan büyük olduğu görülmektedir. Bu da çoklu

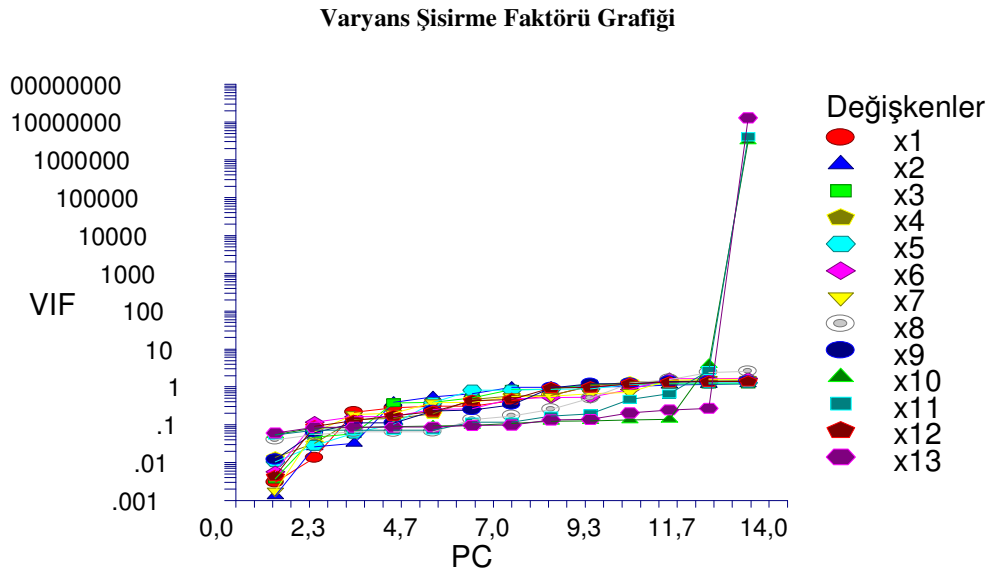
bağlantının varlığına işaretir. Her bir değişkenin diğer değişkenlere olan R^2 değerlerine bakıldığında x_{10} , x_{11} ve x_{13} değişkenleri için bu değer yüksektir. Bağımsız değişkenlerin birbiri ile ilişkili olduğunu göstermektedir. Ayrıca $1-R^2$ olarak hesaplanan tolerans değeri çoklu bağlantı durumunda küçülecektir. Yine x_{10} , x_{11} ve x_{13} değişkenleri için 0 dır. Bu da çoklu bağlantının bir göstergesidir.

Tablo 3.7. Korelasyon Özdeğerleri

No	Özdeğer	Göreceli Yüzde	Birikimsel Yüzde	Koşul sayısı
1	3.833	29.490	29.490	1.000
2	1.992	15.330	44.810	1.924
3	1.593	12.260	57.070	2.410
4	1.130	8.690	65.760	3.390
5	1.014	7.800	73.560	3.780
6	0.723	5.560	79.130	5.300
7	0.700	5.390	84.510	5.480
8	0.635	4.890	89.400	6.030
9	0.479	3.690	93.090	7.990
10	0.411	3.160	96.250	7.990
11	0.341	2.630	98.880	11.230
12	0.145	1.120	100	26.290
13	0	0	100	78226717.890

Tablo3.7' de özdeğerler, göreceli yüzdeler, birikimli yüzdeler ve koşul sayısı sunulmuştur. Özdeğerler, korelasyon matrisinin özdeğerleridir ve koşul sayısı bu değerler arasında en büyük özdeğerin her bir açıklayıcı değişkenin özdeğerine bölünmesi ile hesaplanmıştır. Öz değerlerin sıfıra yakın olması çoklu bağlantının göstergesidir. Ayrıca koşul sayısı çoklu bağlantının varlığını belirlemek için kullanılan yaygın bir yöntemdir.

Koşul sayısının 30 dan büyük olması genel olarak bir çoklu doğrusal bağlantının olduğunu gösterir. Koşul sayısının 100 ile 1000 arasında olması çoklu doğrusal bağlantının güçlü olduğunu 1000 den büyük olması durumunda ise çoklu doğrusal bağlantının ciddi boyutlarda olduğunu gösterir. Tabloda görüldüğü gibi 13. Koşul sayısı 1000 den büyüktür. Buda çoklu doğrusal bağlantının varlığını ve ciddi bir problem olduğunu gösterir.



Şekil 4. VIF Grafiği

Şekil 4'de temel bileşenlerin VIF değerleri üzerindeki etkisi görülmektedir. Uygun temel bileşen sayısı seçilirse tüm VIF değerlerinin 10'dan küçük olması gerekir.

Tablo 3.8. Standartlaştırılmış Temel Bileşenler Regresyon Katsayıları

PC	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
1	0.023	0.014	-0.023	-0.054	-0.041	0.031	0.017	0.078	0.043	0.091	0.089	-0.028	0.096
2	0.068	0.068	-0.093	-0.119	-0.091	0.148	0.100	0.032	0.116	0.027	0.035	-0.135	0.033
3	0.049	0.071	-0.091	-0.107	-0.084	0.157	0.115	0.035	0.126	0.028	0.036	-0.127	0.034
4	0.063	0.116	-0.050	-0.103	-0.070	0.161	0.111	0.034	0.125	0.033	0.036	-0.138	0.036
5	0.065	0.114	-0.051	-0.102	-0.064	0.163	0.107	0.035	0.129	0.033	0.036	-0.140	0.037
6	0.063	0.119	-0.057	-0.098	-0.065	0.160	0.104	0.032	0.130	0.034	0.036	-0.142	0.037
7	0.070	0.129	-0.071	-0.112	-0.045	0.165	0.110	0.024	0.119	0.032	0.044	-0.123	0.040
8	0.099	0.141	-0.070	-0.119	-0.060	0.146	0.098	0.044	0.182	0.023	0.027	-0.068	0.026
9	0.210	0.155	-0.142	-0.002	-0.030	0.138	0.172	0.199	0.103	0.037	-0.036	-0.084	-0.002
10	0.188	0.159	-0.144	-0.036	-0.007	0.115	0.174	0.223	0.107	0.040	-0.055	-0.100	-0.011
11	0.157	0.181	-0.154	-0.046	-0.031	0.172	0.118	0.255	0.078	0.046	-0.081	-0.077	-0.022
12	0.145	0.190	-0.152	-0.042	-0.029	0.169	0.098	0.313	0.078	-0.072	0.005	-0.093	-0.032
13	0.153	0.165	-0.144	-0.027	-0.036	0.188	0.101	0.274	0.090	169.803	192.772	-0.103	-343.232

Tablo3.8’de PC=1 den başlayarak standartlaştırılmış temel bileşen regresyon katsayıları verilmektedir. Bilindiği üzere çoklu bağlantı problemi EKK kestirimlerini işaret ve büyüklük olarak etkileyebilmektedir. Böyle bir durum söz konusu ise bu problemi ortadan kaldıracabilecek bir PC değerinin seçilmesi gerekmektedir.

Tablo 3.9. VIF deęerleri

PC	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
1	0.003	0.001	0.003	0.018	0.010	0.006	0.001	0.038	0.011	0.052	0.050	0.004	0.057
2	0.018	0.022	0.040	0.049	0.029	0.107	0.053	0.055	0.052	0.083	0.072	0.091	0.086
3	0.230	0.031	0.043	0.129	0.057	0.158	0.196	0.062	0.110	0.084	0.072	0.126	0.087
4	0.273	0.447	0.388	0.131	0.099	0.161	0.201	0.062	0.110	0.088	0.072	0.150	0.087
5	0.299	0.534	0.396	0.161	0.472	0.255	0.386	0.062	0.241	0.092	0.072	0.196	0.088
6	0.354	0.886	0.819	0.303	0.475	0.421	0.486	0.116	0.267	0.101	0.078	0.236	0.096
7	0.401	0.973	0.996	0.480	0.813	0.445	0.518	0.174	0.369	0.106	0.132	0.555	0.105
8	0.544	0.995	0.996	0.488	0.850	0.504	0.543	0.239	0.999	0.119	0.181	1.044	0.138
9	0.894	1.001	1.142	0.877	0.876	0.505	0.698	0.918	1.177	0.125	0.296	1.052	0.161
10	1.201	1.008	1.146	1.561	1.188	0.846	0.702	1.277	1.186	0.131	0.505	1.196	0.210
11	1.438	1.130	1.170	1.584	1.327	1.663	1.485	1.525	1.389	0.141	0.669	1.324	0.242
12	1.476	1.153	1.171	1.588	1.328	1.665	1.588	2.405	1.389	3.867	2.659	1.389	0.268
13	1.484	1.223	1.179	1.611	1.333	1.706	1.588	2.575	1.4055	3204156.888	4125897.939	1.401	13078203.449

PC=1 den bařlanarak PC'ler iin VIF deęerleri verilmektedir. Bu tablo VIF grafięinde gosterilen deęerlerin tablosudur. oklu baęlantının giderilmesinde en buyk VIF deęerini lut olarak alındıęında tum deęiřkenlerin VIF deęerinin 10 deęerinin altında olduęu bir PC deęeri seilebilir.

Tablo 3.10. PC Analiz Tablosu

PC	R^2	RMSE	$B' B$	Ortalama VIF	En Büyük VIF
1	0.160	3.096	0.042	0.020	0.057
2	0.294	2.839	0.108	0.058	0.107
3	0.296	2.835	0.110	0.107	0.230
4	0.300	2.826	0.114	0.175	0.447
5	0.301	2.825	0.114	0.250	0.534
6	0.301	2.825	0.114	0.357	0.886
7	0.302	2.823	0.116	0.467	0.996
8	0.308	2.810	0.125	0.588	1.044
9	0.343	2.738	0.199	0.748	1.177
10	0.345	2.735	0.203	0.935	1.561
11	0.349	2.726	0.215	1.160	1.663
12	0.353	2.718	0.241	1.688	3.867
13	0.362	2.699	183803.349	1569867.214	13078203.449

PC nin seçimine bağlı olarak bazı istatistiklerde meydana gelen değişimleri göstermektedir. R^2 değerini maksimize eden ve RMSE değerini minimize eden çözüm EKK' dir. $B' B$ standartlaştırılmış regresyon katsayılarının kareler toplamıdır. PC değerinin seçimine göre bu değerlerin durağanlaşması gerekmektedir. Ortalama VIF değeri her PC değerine karşılık gelen VIF değerinin en büyüğünü verir. Uygun VIF değerinde bu değer 10 dan küçük olur.

Temel bileşen sayısının seçiminde özdeğerler dikkate alınır, özdeğeri 1'den büyük olan bileşenlerin sayısı temel bileşen sayısı olarak seçilebilir. PC=5 değeri alınabilir.

Tablo 3.11. PC=5 için Temel Bileşenler ve EKK'nın Karşılaştırılması

Bağımsız Değ.	Bileşen Kat.	EKK kat.	Std'miş Bileşen	Std'miş EKK	Bileşen Std. Hata	EKK Std. Hata
sabit	23.309	22.611				
λ	0.018	0.043	0.065	0.153	0.012	0.025
x_2	0.016	0.024	0.114	0.165	0.008	0.012
x_3	-0.022	-0.063	-0.051	-0.144	0.021	0.035
x_4	-10.447	-2.853	-0.102	-0.027	3.222	9.739
x_5	-9.720	-5.471	-0.064	-0.036	8.107	13.011
x_6	0.141	0.162	0.163	0.188	0.034	0.084
x_7	0.107	0.100	0.107	0.101	0.048	0.094
x_8	0.096	0.757	0.035	0.274	0.054	0.333
x_9	0.143	0.100	0.129	0.090	0.043	0.099
x_{10}	0.081	41.127	0.033	169.803	0.005	32.576
x_{11}	0.015	82.284	0.036	192.772	0.009	65.145
x_{12}	-0.104	-0.076	-0.140	-0.103	0.025	0.065
x_{13}	0.013	123.428	0.037	343.232	0.008	97.714
R^2	0.30	0.36				
RMSE	2.825	2.699				

PC=5 değerine göre Bileşen katsayıları, EKK katsayıları, Standartlaştırılmış bileşen katsayıların ve standartlaştırılmış EKK katsayıları, standart hataları, R^2 ve RMSE değerleri tabloda görülmektedir. Temel bileşenler regresyonu çoklu bağlantı problemini ortadan kaldırdığı için bu yöntemle elde edilen kestirimlerin standart hatalarının EKK regresyonu ile elde edilen kestirimlerden daha düşük çıkması gerekmektedir. Tablodan da görüleceği üzere özellikle çoklu bağlantıya sebep olan x_{10} , x_{11} ve x_{13} değişkenlerin kestirimlerinde standart hataları çok ciddi oranda azalmıştır. Ayrıca analiz sonucunda R^2 , EKK için 0.36, temel bileşenler için 0.30 olarak hesaplanmıştır.

Tablo3.12. Temel Bileşenler Regresyon Katsayıları

Bağımsız Değişkenler	Regresyon Katsayıları	Standart Hata	Std'miş Reg. Katsayıları	VIF	p değeri
Sabit	23.309				
x_1	0.018	0.012	0.065	0.299	0.132
x_2	0.016	0.008	0.114	0.534	0.047*
x_3	-0.022	0.021	-0.051	0.396	0.299
x_4	-10.447	3.222	-0.102	0.161	0.001*
x_5	-9.720	8.107	-0.064	0.472	0.232
x_6	0.141	0.034	0.163	0.255	0.000*
x_7	0.107	0.048	0.107	0.386	0.029*
x_8	0.096	0.054	0.035	0.062	0.000*
x_9	0.143	0.043	0.129	0.241	0.001*
x_{10}	0.008	0.005	0.033	0.092	0.157
x_{11}	0.015	0.009	0.036	0.072	0.000*
x_{12}	-0.104	0.025	-0.140	0.196	0.000*
x_{13}	0.013	0.008	0.037	0.088	0.115

Tablo3.12'de PC=5 değeri için Temel Bileşenler Regresyonunun detaylı çözümü verilmektedir. Söz konusu değişkenler için regresyon denklemimiz;

$$Y=23.3090+0.01664 x_2 -10.4470 x_4 +0.1413 x_6 +0.1071 x_7 +0.0966 x_8 +0.1439 x_9 +0.0154 x_{11} -0.1042 x_{12} \text{ dir.}$$

Tablo 3.13. PC=5 için ANOVA Tablosu

Değ. Kay.	Sd	HKT	HKO	F-Oranı	p
sabit	1	165457.100	165457.100		
model	13	388.518	29.886	3.742	<0.00005
Hata	113	902.347	7.985		
Toplam	126	1290.866	10.244		
Bağ. Değ. Ort.	36.094				
RMSE	2.825				
R^2	0.30				
Değişim Kats.	0.078				

Tabloda PC=5 için Temel Bileşenler Regresyon modelinin önemliliğini test eden varyans analizine ilişkin bilgiler sunulmuştur. Tablodan görüldüğü gibi $p < 0.05$ olduğu için Temel Bileşenler Regresyon modeli anlamlıdır. Bir başka ifadeyle sol atriyum çapını istatistiksel olarak önemli değişkenlerle açıklanabileceği söylenebilir. O halde incelenen veriler için sol atriyum çapı %30 oranında tahmin edilebilir.

PLSR Analizi Minitab15 ve PASW18 programları kullanılarak yapılmıştır. Yapılan analiz sonucunda X matrisi sıfır oluncaya kadar sürdürülmesiyle 12 değişkenin açıklayıcı ve yanıt değişkenindeki değişimin tamamını açıkladığı görülmüştür. Açıklayıcı ve yanıt değişkenlerine ait ağırlık ve yük vektörleri tablolar halinde gösterilmiştir.

Tablo 4.2. Bileşenlere Ait Ağırlık Değerleri

Bağımsız Değişkenler	Bileşenler											
	1	2	3	4	5	6	7	8	9	10	11	12
x_1	0.231	0.262	0.320	-0.016	-0.586	-0.033	0.028	0.314	0.206	-0.187	-0.261	-0.267
x_2	0.264	0.399	0.337	-0.378	0.040	0.305	0.050	-0.169	-0.287	0.326	-0.139	0.372
x_3	-0.244	-0.234	-0.158	-0.538	0.062	0.264	-0.148	0.211	-0.247	-0.321	0.198	-0.355
x_4	-0.298	0.016	0.264	-0.064	0.129	-0.205	-0.077	0.346	0.243	-0.376	-0.519	0.367
x_5	-0.216	0.045	0.227	-0.264	-0.013	-0.024	0.043	-0.091	0.755	0.382	0.266	-0.215
x_6	0.387	0.362	-0.049	-0.002	0.298	0.093	-0.131	-0.414	-0.014	-0.375	-0.454	-0.430
x_7	0.268	0.294	-0.019	-0.163	-0.087	-0.583	-0.095	0.499	0.030	0.392	0.066	-0.304
x_8	0.338	0.102	0.812	0.542	0.506	0.353	0.125	0.095	-0.062	-0.120	0.314	-0.065
x_9	0.333	0.161	-0.149	-0.356	-0.186	-0.206	0.419	0.051	0.077	-0.410	0.359	0.299
x_{10}	0.227	-0.376	-0.056	-0.361	-0.572	-0.880	-0.841	-0.403	-0.089	-0.036	-0.008	0.091
x_{11}	0.207	-0.433	-0.199	0.005	0.385	0.641	0.680	0.400	0.273	0.158	-0.254	-0.052
x_{12}	-0.299	-0.199	0.239	-0.120	-0.199	-0.323	0.359	-0.304	-0.454	-0.087	-0.231	-0.322
x_{13}	0.229	-0.429	-0.140	-0.176	-0.067	-0.075	-0.035	0.025	0.110	0.071	-0.147	0.016
Y	0.300	0.150	0.177	0.043	0.036	0.024	0.018	0.006	0.001	0.000	0.000	0.000

Tablo 4.2’de açıklayıcı değişkenlere ve yanıt değişkenine ait ağırlık değerleri verilmektedir. Açıklayıcı değişkenler için ağırlık değeri ister pozitif ister negatif olsun her bir değişkenin bileşene katkısını göstermektedir. Yanıt değişkeni için ağırlık değeri ise, her bir bileşenin yanıt değişkenini modellemede katkısını göstermektedir. Başka bir deyişle yanıt değişkeni için bileşenin ağırlık değeri ne kadar büyükse modelleme katkısı o kadar büyük denebilir.

Tablo 4.3. Bileşenlere Ait Yük Değerleri

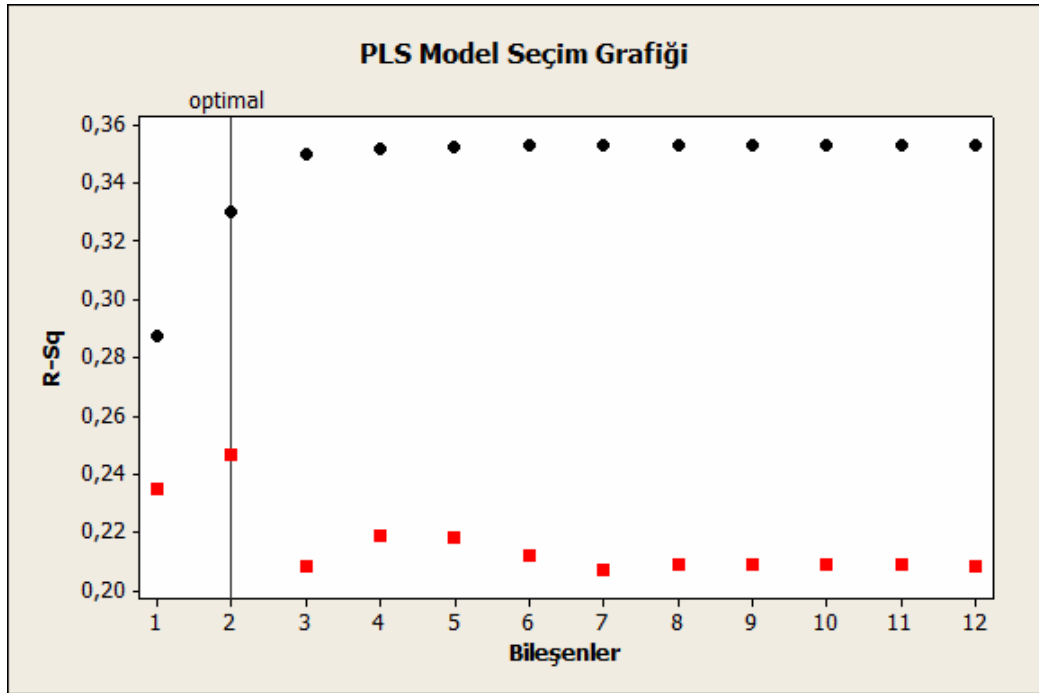
Değişkenler	Bileşenler											
	1	2	3	4	5	6	7	8	9	10	11	12
x_1	0.172	0.108	0.258	0.162	-1.016	0.466	-0.049	0.272	0.149	-0.183	-0.211	-0.255
x_2	0.168	0.248	0.344	-0.600	-0.016	0.326	-0.052	-0.108	-0.310	0.428	-0.224	0.378
x_3	-0.196	-0.134	0.095	-0.646	0.140	0.358	-0.366	0.362	-0.262	-0.325	0.281	-0.364
x_4	-0.330	0.005	0.308	-0.231	0.447	-0.336	-0.085	0.328	0.215	-0.334	-0.459	0.390
x_5	-0.249	0.028	0.322	-0.396	0.127	-0.041	0.088	-0.365	0.739	0.188	0.196	-0.227
x_6	0.313	0.319	-0.194	-0.133	0.453	-0.079	-0.065	-0.414	0.200	-0.294	-0.358	-0.410
x_7	0.204	0.265	-0.081	-0.153	0.444	-0.614	0.046	0.574	-0.215	0.389	-0.002	-0.307
x_8	0.338	-0.301	0.682	0.154	0.328	-0.068	-0.077	0.087	-0.071	-0.178	0.342	-0.079
x_9	0.314	0.143	-0.109	-0.301	0.024	-0.314	0.550	-0.099	0.147	-0.519	0.432	0.283

Tablo 4.3'ün devamı...

x_{10}	0.361	-0.480	0.204	-0.149	-0.065	-0.169	-0.337	-0.158	0.049	-0.018	-0.005	0.091
x_{11}	0.356	-0.483	-0.054	-0.117	0.118	0.116	0.286	0.143	0.120	0.161	-0.285	-0.041
x_{12}	-0.266	-0.233	0.383	-0.131	-0.011	-0.421	0.657	-0.298	-0.354	0.050	-0.199	-0.311
x_{13}	0.379	-0.509	0.070	-0.139	0.034	-0.018	-0.006	0.002	0.091	0.082	-0.163	0.022
Y	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Tablo 4.3'de yük değerleri gösterilmiştir. Yük değerleri bileşenlerin açıklayıcı değişkeni açıklama miktarını göstermektedir.

Bileşenlere ait ağırlık ve yük vektörleri bulunduktan sonra ikinci adım regresyon analizidir. Regresyon analizinde hangi bileşenlerin modelde kalması gerektiğini belirlemek için tekli çapraz geçerlilik yöntemi uygulanmıştır. Bu yöntem sırası ile her bir gözlemin modelden çıkararak geri kalan gözlemler ile model kurmaya dayanmaktadır. Çıkarılan her gözlem tekrar yerine konmaktadır. Tahmini hata kareler toplam değeri (PRESS) hesaplanarak en küçük PRESS değerine sahip olan bileşen sayısı ideal bileşen sayısı olarak belirlenmektedir.



Şekil 5. PLS Model Seçim Grafiği

Şekil5 PLSR analizi için ideal modelin kaç bileşenli olacağını göstermektedir. Şekilde görüldüğü gibi ideal bileşen sayısının 2 olduğu söylenebilir.

Tablo 4.4. PLSR Model Seçimi ve Geçerlilik Tablosu

Bileşenler	X Varyans	Hata kareler toplamı	R^2	PRESS
1	0.267	920.167	0.287	987.980
2	0.437	864.455	0.330	972.750
3		838.787	0.350	1022.310
4		836.710	0.351	1008.750
5		835.863	0.352	1008.990
6		835.367	0.352	1017.310
7		835.173	0.353	1023.420
8		835.152	0.353	1020.910
9		835.150	0.353	1021.490
10		835.150	0.353	1021.560
11		835.150	0.353	1021.580
12		835.150	0.353	1021.830

Tablo4.4'te bağımsız değişkenlerdeki varyans değeri, bileşenlere ait hata kareler toplamı, R^2 ve PRESS değerleri verilmiştir. Görüldüğü gibi 2 bileşenli PLSR modeli bağımsız değişken varyansının yaklaşık %44 ünü açıklamaktadır. PRESS değeri en küçük değeri almıştır. R^2 ise 0.33 olarak hesaplanmıştır.

Tablo 4.5. Regresyon Katsayıları Tablosu

Bağımsız Değişkenler	Regresyon Katsayıları	Std'miş Reg. Katsayıları	VIP
Sabit	23.206	0.000	
x_1	0.030	0.108	0.831*
x_2	0.020	0.138	0.951*
x_3	-0.047	-0.108	0.882*
x_4	-8.886	-0.087	1.074**
x_5	-8.729	-0.058	0.779
x_6	0.147	0.170	1.395**
x_7	0.124	0.124	0.966*
x_8	0.322	0.116	1.218**
x_9	0.138	0.123	1.199**
x_{10}	0.002	0.011	0.820*
x_{11}	-0.001	-0.002	0.746
x_{12}	-0.088	-0.119	1.077**
x_{13}	0.001	0.004	0.825*

Tablo4.5'te regresyon katsayıları, standartlaştırılmış regresyon katsayıları ve VIP değerleri gösterilmektedir. VIP değerleri her bir bağımsız değişkenin bağımlı değişkeni açıklamadaki katkısını gösterir. Eğer VIP değeri 0.8'den küçükse o değişkenin bağımlı değişkeni açıklamaya katkısı az ve modelden çıkarılmaya adaydır denilebilir[1].

Tablo 4.6. PLSR(2) İçin ANOVA Tablosu

Kaynak	Sd	HKT	HKO	F	P
Regresyon	2	426.410	213.206	30.580	0.0001
Hata	124	864.450	6.971		
Toplam	126	1290.870			

Tabloda PLS(2) için Kısmi En Küçük Kareler Regresyonu modelinin önemliliğini test eden varyans analizine ilişkin bilgiler sunulmuştur. Tablodan görüldüğü gibi $p < 0.05$ olduğu için Kısmi En Küçük Kareler Regresyon modeli anlamlıdır.

YSA'ya 127 bireye ait veri kümesinden örnekler rastgele seçilip verilerek, eğitim doğrulama ve test yapılmıştır. Öğrenme algoritması olarak Levenberg Marquardt algoritması kullanılarak en iyi sonuç elde edilmeye çalışılmıştır. YSA için matlab programından yararlanılmıştır.

Tablo5.1. YSA Modellerine İlişkin Bilgiler ve Sonuçları

YSA	Katmandaki Nöron Sayısı			Örnek Büyüklükleri			Ö.A	SONUÇ	
	Girdi Katmanı	Ara Katman	Çıktı katmanı	Eğitim	Doğrulama	Test		HKO	R^2
1	13	15	1	89	19	19	LM	6.717	0.40
2	13	18	1	89	19	19	LM	6.612	0.38
3	13	20	1	89	19	19	LM	6.162	0.41*
4	13	24	1	89	19	19	LM	7.345	0.42
5	13	15	1	77	25	25	LM	6.987	0.34
6	13	18	1	77	25	25	LM	7.761	0.42
7	13	20	1	77	25	25	LM	7.027	0.36
8	13	22	1	77	25	25	LM	7.230	0.39
9	13	15	1	63	32	32	LM	7.025	0.32
10	13	20	1	63	32	32	LM	7.382	0.35

Ara katmandaki nöron sayısı ve eğitim, test, doğrulama değerleri değiştirilerek YSA modelleri oluşturulmuş, en iyi 10 sonuç tabloda gösterilmiştir. Sonuçlar değerlendirildiğinde, HKO değerinin en düşük olduğu değer dikkate alınarak seçim yapılırsa 3. Model en başarılı model olarak belirlenebilir.

4. SONUÇ

Bu çalışmada açıklayıcı değişkenler arasındaki çoklu doğrusal bağlantı olması durumunda kullanılan PCR, PLSR regresyonu ve YSA ile sol atriyum çapının tahmini yapılarak elde edilen sonuçlar karşılaştırılmıştır.

Tablo 6.1. Tahminlerin Karşılaştırılması

	PCR	PLSR	YSA
RMSE	2.825	2.640	2.482
R^2	0.30	0.33	0.41

Tablo6.1’de PCR, PLSR, YSA modellerine ait RMSE ve R^2 değerleri görülmektedir. Tahminlerin birbirine üstünlüklerini belirlemede R^2 belirlilik katsayısı ve RMSE değerleri kullanılmıştır. Sonuçlar değerlendirildiğinde YSA ile yapılan tahminin PCR ve PLSR’ye göre RMSE değerinin daha düşük olduğu ve R^2 belirlilik katsayısının daha yüksek olduğu görülerek YSA’nın sol atriyum çapının tahmininde daha iyi sonuç verdiği belirlenmiştir.

5. EKLER

EK Tablo 1.

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	Y
1	62	81	39	0.02	0.01	32	43	9	31	140	90	45	106.67	39
2	66	111	47	0.05	0.02	26	41	11	32	150	90	60	110.00	45
3	64	96	43	0.02	0.08	33	46	7	30	140	90	55	106.67	36
4	59	74	43	0.05	0.06	31	44	11	33	160	100	50	120.00	33
5	55	59	53	0.03	0.02	26	41	8	31	140	90	60	106.67	31
6	60	104	53	0.07	0.03	34	48	9	33	140	90	50	106.67	39
7	54	111	43	0.04	0.12	28	44	8	36	130	80	55	96.67	36
8	60	104	43	0.02	0.04	28	48	9	34	140	90	50	106.67	36
9	26	74	50	0.09	0.04	29	47	8	32	140	80	60	100.00	32
10	40	104	40	0.04	0.03	27	47	8	34	130	90	60	103.33	33
11	60	74	53	0.03	0.04	45	57	11	36	150	100	60	116.67	43
12	56	74	46	0.03	0.04	30	41	8	34	130	80	45	96.67	35
13	51	71	56	0.05	0.06	29	45	9	32	130	80	55	96.67	35
14	56	104	43	0.09	0.03	33	46	9	34	140	90	55	106.67	37
15	53	67	51	0.01	0.02	29	45	11	35	150	90	55	110.00	38
16	20	59	53	0.13	0.04	25	47	10	32	140	90	65	106.67	33
17	33	96	50	0.03	0.03	28	47	10	33	150	100	50	116.67	37

EK Tablo1 devamı...

18	60	81	47	0.08	0.04	24	44	8	25	120	80	60	93.33	40
19	54	74	43	0.04	0.03	26	43	10	31	150	90	55	110.00	32
20	51	74	45	0.02	0.03	28	45	11	32	150	90	55	110.00	38
21	45	21	43	0.04	0.03	25	42	13	28	170	100	60	123.33	35
22	40	96	56	0.11	0.05	28	46	10	36	160	90	60	113.33	32
23	65	104	44	0.02	0.02	27	43	9	34	130	90	60	103.33	42
24	55	59	49	0.03	0.03	28	47	8	31	120	80	55	93.33	38
25	58	104	43	0.05	0.04	29	45	9	32	130	90	55	103.33	43
26	54	96	39	0.04	0.04	28	44	11	36	160	90	55	113.33	41
27	62	74	47	0.02	0.02	28	47	9	36	140	90	55	106.67	39
28	43	74	43	0.06	0.09	26	42	9	32	120	80	55	93.33	36
29	56	81	40	0.06	0.05	30	47	9	36	130	90	60	103.33	36
30	59	89	43	0.06	0.05	26	43	9	41	140	80	55	100.00	35
31	39	89	52	0.10	0.05	24	43	9	32	140	90	60	106.67	37
32	63	67	41	0.03	0.02	24	43	8	31	110	80	60	90.00	33
33	62	81	53	0.04	0.03	23	40	8	28	120	80	55	93.33	33
34	62	81	48	0.05	0.02	25	44	11	35	160	100	60	120.00	40
35	64	104	47	0.03	0.02	27	44	10	39	150	90	55	110.00	42
36	62	59	39	0.04	0.03	27	43	8	32	130	80	60	96.67	39

EK Tablo1 devamı...

37	55	74	44	0.03	0.04	32	49	11	36	150	100	55	116.67	33
38	57	89	46	0.03	0.02	28	42	11	32	150	90	50	110.00	40
39	65	81	45	0.05	0.03	28	42	9	32	130	80	55	96.67	33
40	45	59	46	0.05	0.04	28	44	10	33	140	90	65	106.67	32
41	61	89	35	0.04	0.02	28	43	11	34	160	110	50	126.67	42
42	47	81	70	0.12	0.09	23	39	8	30	130	80	60	96.67	31
43	39	67	50	0.04	0.07	24	41	8	32	140	90	60	106.67	32
44	55	59	42	0.07	0.05	26	44	10	33	150	90	55	110.00	34
45	60	67	44	0.06	0.03	28	44	8	30	140	90	55	106.67	32
46	67	67	34	0.04	0.03	27	41	11	36	170	90	55	116.67	37
47	70	111	62	0.05	0.03	26	42	9	33	140	90	55	106.67	36
48	70	59	48	0.02	0.02	32	48	10	32	140	90	55	106.67	38
49	69	67	47	0.02	0.02	28	40	9	34	130	80	55	96.67	32
50	66	89	52	0.02	0.02	24	43	8	33	120	80	55	93.33	41
51	75	96	36	0.02	0.02	29	41	9	29	140	90	55	106.67	34
52	45	81	42	0.04	0.02	27	44	8	34	140	90	55	106.67	35
53	50	67	47	0.08	0.03	23	42	8	28	130	80	65	96.67	31
54	60	224	68	0.03	0.03	29	46	9	33	150	90	55	110.00	36
55	75	59	45	0.02	0.02	26	40	9	35	140	90	55	106.67	37

EK Tablo1 devamı...

56	46	59	55	0.17	0.12	29	44	9	35	130	80	60	96.67	35
57	53	67	44	0.07	0.08	24	43	9	30	130	70	60	90.00	37
58	51	96	53	0.03	0.04	28	46	9	31	140	80	55	100.00	36
59	40	104	40	0.10	0.06	21	40	8	31	130	90	60	103.33	32
60	37	111	32	0.03	0.03	38	51	10	38	150	90	50	110.00	43
61	53	96	50	0.04	0.03	25	42	9	27	140	80	55	100.00	42
62	54	96	42	0.09	0.05	26	40	8	31	130	80	60	96.67	35
63	20	89	43	0.11	0.02	26	48	9	36	120	80	65	93.33	34
64	52	111	44	0.03	0.04	31	46	9	35	130	80	55	96.67	42
65	40	81	55	0.11	0.07	27	46	8	29	120	80	65	93.33	33
66	39	74	52	0.06	0.06	27	44	7	30	130	90	65	103.33	36
67	32	65	43	0.06	0.02	30	45	8	34	130	80	55	96.67	38
68	49	81	49	0.11	0.04	26	42	8	28	130	80	60	96.67	33
69	50	89	66	0.04	0.03	26	42	11	34	160	90	55	113.33	36
70	51	57	43	0.03	0.03	25	39	9	30	160	100	55	120.00	39
71	62	52	42	0.02	0.02	32	45	8	35	140	90	45	106.67	35
72	48	81	61	0.05	0.03	29	42	8	30	110	70	55	83.33	33
73	65	74	39	0.03	0.02	26	43	10	38	160	110	60	126.67	37
74	58	51	36	0.03	0.05	29	45	9	33	120	80	60	93.33	37

EK Tablo1 devamı...

75	46	89	51	0.03	0.01	27	47	12	35	150	90	65	110.00	34
76	60	59	58	0.05	0.06	30	47	9	34	130	80	55	96.67	38
77	48	67	41	0.10	0.05	25	44	8	31	140	90	65	106.67	35
78	49	67	38	0.04	0.02	32	48	8	32	130	80	55	96.67	33
79	53	67	47	0.03	0.02	26	42	10	33	170	100	60	123.33	32
80	57	59	41	0.04	0.03	31	43	9	40	160	100	55	120.00	39
81	57	67	48	0.01	0.04	30	46	9	32	150	90	50	110.00	41
82	55	67	49	0.05	0.09	45	28	9	41	140	90	55	106.67	36
83	54	74	44	0.04	0.03	27	41	8	36	140	80	60	100.00	36
84	58	110	45	0.04	0.02	31	48	10	35	150	90	55	110.00	36
85	50	81	44	0.03	0.02	27	43	9	34	150	90	50	110.00	33
86	52	81	47	0.01	0.02	24	45	8	32	140	90	60	106.67	32
87	56	74	39	0.05	0.05	24	41	9	29	140	80	55	100.00	32
88	44	74	41	0.05	0.05	26	42	10	28	150	90	55	110.00	39
89	42	96	48	0.02	0.06	30	47	10	37	140	90	60	106.67	38
90	60	81	49	0.04	0.03	27	43	9	34	130	90	55	103.33	34
91	65	118	44	0.08	0.02	27	45	9	35	140	80	50	100.00	36
92	65	111	45	0.06	0.02	30	42	7	29	140	90	55	106.67	35
93	59	111	47	0.01	0.03	33	49	9	36	140	90	50	106.67	39

EK Tablo1 devamı...

94	50	111	44	0.04	0.02	32	45	10	31	160	100	55	120.00	37
95	56	98	45	0.04	0.02	30	44	8	29	120	80	55	93.33	34
96	40	67	45	0.02	0.01	27	44	9	32	140	90	60	106.67	35
97	47	67	50	0.04	0.03	28	44	10	32	140	90	65	106.67	36
98	57	96	52	0.04	0.03	32	45	10	32	150	90	60	110.00	39
99	49	81	62	0.03	0.06	25	43	11	29	170	100	65	123.33	35
100	60	89	43	0.03	0.04	25	40	11	32	160	100	60	120.00	34
101	55	57	54	0.03	0.02	27	41	8	31	140	90	55	106.67	33
102	56	103	43	0.08	0.02	32	45	9	34	140	90	50	106.67	37
103	42	73	42	0.05	0.09	25	42	9	32	150	100	60	116.67	34
104	35	111	32	0.03	0.03	38	51	10	38	150	90	55	110.00	42
105	24	89	43	0.11	0.02	26	48	9	36	130	80	55	96.67	34
106	27	74	50	0.09	0.04	29	47	8	32	130	80	60	96.67	34
107	46	81	70	0.12	0.09	23	39	8	30	140	90	60	106.67	31
108	31	65	43	0.06	0.02	30	45	8	34	140	90	55	106.67	38
109	63	96	43	0.02	0.08	33	46	7	30	120	80	55	93.33	34
110	51	81	47	0.01	0.02	24	45	8	32	140	90	60	106.67	30
111	50	67	51	0.01	0.02	29	45	11	35	150	100	55	116.67	38
112	61	89	35	0.04	0.02	28	43	11	34	160	100	60	120.00	38

EK Tablo1 devamı...

113	49	89	66	0.04	0.03	26	42	11	34	160	90	55	113.33	36
114	72	59	48	0.02	0.02	32	48	10	32	140	90	55	106.67	36
115	44	74	41	0.05	0.05	26	42	10	28	150	90	60	110.00	32
116	40	104	40	0.04	0.03	27	47	8	34	130	90	55	103.33	37
117	64	74	39	0.03	0.02	26	43	10	38	150	90	55	110.00	40
118	43	59	55	0.17	0.02	29	44	9	35	130	80	50	96.67	35
119	32	96	50	0.03	0.03	28	47	10	33	140	90	55	106.67	35
120	37	111	32	0.03	0.03	38	51	10	38	150	90	55	110.00	42
121	49	67	41	0.10	0.05	25	44	8	31	120	80	55	93.33	34
122	51	96	53	0.03	0.04	28	46	9	31	130	90	60	103.33	33
123	33	65	43	0.06	0.02	30	45	8	34	130	80	55	96.67	36
124	52	57	43	0.03	0.03	25	39	9	30	140	90	50	106.67	34
125	67	111	47	0.05	0.02	26	41	11	32	170	100	50	123.33	40
126	34	96	50	0.03	0.03	28	47	10	33	170	110	55	130.00	37
127	65	104	44	0.02	0.02	27	43	9	34	140	90	55	106.67	38

6. KAYNAKLAR

- [1]. **Abudu, S.**, 2010. Modeling of Daily Pan Evapoation Using Partial Least Squares Regression.
- [2]. **Albayrak, A. S.**, 2005. Çoklu Doğrusal Bağlantı halinde EKK Tekniğinin Alternatifi Yanlı Tahmin Teknikleri ve Bir Uygulama. ZKÜ Sosyal Bilimler Dergisi, 1.1.
- [3]. **Bulut, E.**, 2009. Kısmi En Küçük Kareler Regresyon Yöntemi Algoritmalarından Nipals ve PLS- Kernel Algoritmalarının Karşılaştırılması ve Bir uygulama
- [4]. **Chapter 9.** Biased Regression.
<http://www.plantsciences.ucdavis.edu/agr206/AGR206files/AGR206Lectures/AGR206Ch09PCR.pdf>
- [5]. **Collins, B.**, 2010. Partial Least Squares Regression.
- [6]. **D.P.T.**, İllerin ve Bölgelerin Sosyo-Ekonomik Gelişmişlik Sırası ile Araştırılması, Ankara, 2003.
- [7]. **Ekiz, U.**, 1997. Çoklu Bağlantıya Neden Olan Gözlemlerin Belirlenmesi, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- [8]. **Giuseppe, P.**, 2009. Performance of PLS Regression Coefficients in Selecting Variables for Each Response of a Multivariate PLS for Omics-Type Data
- [9]. **Draper, N.**, 1966. Applied Regression Analysis.
- [10]. **İşi, A.**, 2002. Yanlı Tahmin ediciler ve Kombinasyonları, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- [11]. **İpek, O.**, 2000. Ridge Regresyon Üzerine Bir Çalışma, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- [12]. **Karagöz, M.**, 2006. İstatistik Yöntemleri, Ekin Yayıncılık, Bursa.
- [13]. **Karakaş, S.**, 2008. Çoklu Doğrusal Bağlantı Problemi ve Yanlı Regresyon Tahminçileri, Yüksek Lisans Tezi, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü.
- [14]. **Kaşko, Y.**, 2007. Çoklu Bağlantı Durumunda İkili (Binary) Lojistik Regresyon Modelinde Gerçekleşen I. Tip Hata ve Testin Gücü, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- [15]. **Kaya, M.**, 2010. Prostat Spesifik Antijeni (PSA)'nın Değişik Yapay Sinir Ağı Modelleri ile Tahmini, Yüksek Lisans Tezi, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elazığ.

- [16]. **Norliza Binti, A.**, 2006. Comparing Three Methods of Handling Multicollinearity Using Simulation Approach, Universiti Teknologi Malaysia
- [17]. **Onay, İ.**, 2003. Ortalama artelyel Kan Basıncını Etkileyen Faktörlerin Temel Bileşenler Regresyonu İle Belirlenmesi, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü. Ankara.
- [18]. **Ortabaş, N.**, 2001. Principal Components in the Problem of multicollinearity, Dokuz Eylül University, İzmir.
- [19]. **Özkan, K.**, 2009. Toprağın Tarla Kapasitesi Değişiminin Toprak Türüne Göre Temel Bileşenler Regresyonu İle Modellenmesi, SDÜ Orman Fakültesi Dergisi, Sayfa 1-9.
- [20]. **Pamukçu. E.**, 2009. Sistolik Kan Basıncını Etkileyebilecek Faktörlerin Ridge Regresyon Analizi ile İncelenmesi ve Çoklu Bağlantı Problemi, Yüksek Lisans Tezi, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elazığ.
- [21]. **Polat. E.**, 2009. Kısmi En Küçük Kareler Regresyonu, Yüksek Lisans Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- [22]. **Tatlıdil. H.**, 2002. Uygulamalı Çok Değişkenli İstatistiksel Analiz, Ankara.
- [23]. **Tolon, M.**, 2008. Yapay Sinir Ağları ve Regresyon Analizi Karşılaştırılması, Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi (247-259).
- [24]. **Okkan, U.**, 2010. Yiğitler Çayı Günlük Akımlarının Yapay Sinir Ağları ve Regresyon Analizi ile Modellenmesi, Dumlupınar Üniversitesi, Fen Bilimleri Enstitüsü Dergisi (sayı 23)
- [25]. **Yazid, M. Al-Hassan., Mowafaq, M. Al-Kassab.**, 2009. A Monte Carlo Comparison between Ridge and Principal Components Regression Methods.

7. ÖZGEÇMİŞ

1986 yılında Elazığ'da doğan Fatma AŞKIN, orta öğrenimini, 2003 yılında, yüksek öğrenimini ise 2005-2009 yılları arasında Fırat Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümünde tamamlamıştır. 2009 yılında Fırat Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Uygulamalı İstatistik Bilim Dalında yüksek lisans eğitimine başlamıştır ve halen eğitimini sürdürmektedir.